1983

Crataegus Crus-galli L Sensu Lato In Southern Ontario: Phenotypic Variation And Variability In Relation To Reproductive Behavior

Timothy Adam Dickinson

CANADIAN THESES ON MICROFICHE

I.S.B.N.

THESES CANADIENNES SUR MICROFICHE

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

NL-339 (r. 82/08)

Canada

CRATAEGUS CRUS-GALLI L. SENSU LATO IN SOUTHERN ONTARIO: Phenotypic Variation and Variability in relation to Reproductive Behavior

by

Timothy Adam DICKINSON

Department of Plant Sciences

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario

June, 1983

VOLUME ONE

ABSTRACT

Crataegus crus-galli L. sensu lato is one of the most widespread groups of hawthorn taxa in eastern North America. Early twentieth century classifications of this group have recognized a considerable degree of dissection into rather narrowly defined species. However, the status of many of these species has been unclear in view of the widespread belief that apomixis is frequent throughout the entire genus. Crataegus crus-galli s. l. was examined at 11 sites across southern Ontario in order to characterize the pattern of morphological variation present in one part of its range. Cluster analyses that employed a selection of resemblance functions and sorting algorithms confirmed the breakdown of the entire complex sampled at these sites into a total of four morphotypes. Two of these are more or less widely distributed and differ from each other most obviously in phenology and stamen number. The other two are much more restricted in their occurrence in Ontario and differ from each other and from the two more common morphotypes in features of flower, foliage and thorn morphology. The degree of multivariate variability exhibited within individual stands of the two common morphotypes was found to be quite low. However, stands of each morphotype were differentiated from one another by a number of correlated

flower, fruit and foliage characteristics. Only triploid (pollen-infertile or completely male-sterile) and tetraploid (pollen-fertile) individuals of C. crus-galli s.l. have been found so far in Ontario (x = 17). All four morphotypes produce apparently unreduced, aposporous embryo-sacs. Results of pollination experiments showed that three more or less pollen-fertile morphotypes are self-compatible; in all four agamospermy is evidently pseudogamous. Together, these results suggest that stands of a given morphotype of C. crus-galli s. l. frequently consist largely of sibling individuals. This may result from apomixis and self-fertility, together with certain characteristics of Crataegus seed dispersal. Unnecessarily narrow species concepts applied to this group in the past were probably a result of confounding variation among such stands with that among taxa.

Why yes, individuals make history; but without knowing what history it is that they are making.

Paraphrase of Karl Marx in "The 18th Brumaire of Louis Napoleon," Die Revolution, 1852.

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Dr. J. B. Phipps, for giving me the opportunity to study Crataegus crus-galli both in the field and in the laboratory. I am deeply indebted to him for having shared with me his insights into hawthorn systematics. The extensive fieldwork and computing required by this project would not have been possible without access to the financial support of his research grant from the Natural Sciences and Engineering Research Council of Canada. An NSERCC equipment grant to Dr. Phipps in 1980 provided the research microscope necessary for the cytological and embryological studies described in Chapters 9 and 10.

Financial support in the form of a Graduate Scholarship from the Province of Ontario, 1978-1979, and Special University Scholarships from the Faculty of Graduate Studies, U. W. O. are gratefully acknowledged.

I would like to express my sincere thanks to the members of my advisory committee, Dr. R. I. Greyson and Dr. M. A. Maun for their encouragement and helpful suggestions. I am indebted to Dr. L. Orloci and Dr. M. E. Redrupp for their comments on the initial research proposal. Dr. Redrupp also assisted preliminary cytological studies. Dr. D. B. Walden generously loaned the micro-press used in making chromosome preparations, and read and commented on a draft of Chapter 9. I would also like to thank Dr.

R. B. van Huystee for helping me carry out preliminary studies of Crataegus seed protein polymorphism in his laboratory during 1977 and 1978.

I would like to express my gratitude to the entire Department of Plant Sciences, particularly all of its support staff and students. I am struck by the way in which work in this department is made much easier and more enjoyable by the warm and supportive relations prevailing among people here. Within the department I have a special indebtedness to several people who are or have been my colleagues here. For several years I have had the pleasure of sharing a genus with Mr. P. G. Smith, and a great deal of my thinking about hawthorn biology and systematics has developed from our discussions. Throughout my time at U. W. O. Dr. J. Bowles has been both a valued friend and critic of my ideas. She read and commented on much of this thesis in draft form, and skilfully prepared the illustration which appears as Figure 3.1. Dr. J. Podani introduced me to the comparison of classifications and provided the programs used not only in the analyses described in Section 6.3 but also in most of the other cluster analyses as well. Dr. R. Gittins taught me about canonical analysis and helped to develop my thinking about data analysis. I am extremely grateful to each one of these persons for their friendship and help during the course of the work whose results are reported here.

Several other persons have assisted materially with this study. Ms G. deBoer, Ms J. Waite, Mr. R. Hopkins, Ms M. Macleod, Ms N. Power, Ms B. Maloney and Ms M. Shaik provided valuable assistance in the field at different times, 1978-1981. Ms Macleod is responsible for the painstaking determinations of pollen grain number, size and stainability reported in Chapter 10 and Appendix 4. She also prepared the illustrations which appear as Figure 4.2 and Figure 6.17. Ms Shaik is responsible for embedding and sectioning much of the embryological material. Ms Power made most of the observations on the insects associated with hawthorns (Chapter 10, Appendix 5). Ms D. Reimer assisted with the cytological studies during the summer of 1980 and is responsible for the photographs used in Plate 2a-c and Plate 3a, b and e. Mr. P. Fewster generously made available to me his programs for plotting stereo diagrams of ordinations (Fig. 6.1a, 6.2a). Mr. P. F. Ulf-Hansen prepared most of the data on species co-occurrence used in Figure 10.4. I am indebted to Mr. A. Noon not only for his skilful printing of the photographic plates but also for his generosity in giving me the benefit of his advice and equipment. Ms S. Willis retyped several chapters inadvertently lost from magnetic tape.

Much of the study reported here would have been impossible without the generosity of Mr. C. Gawne of Florence, Ontario and Mr. A. Carrothers of Appin, Ontario

in allowing me to work on their properties. Likewise, I wish to thank the Superintendent of the Fort George National Historic Park, Mr. C. Halvorsen and his assistant, Mr. R. Andres for making possible my work at that site.

The photographs in Plate 7 were taken on the microscope of my friend Dr. U. Posluszny, of the Department of Botany and Genetics, Guelph University. Mr. J. G. K. Kockel of the Instruments Division, Nikon Canada Inc. kindly made available the stereomicroscope on which the photographs in Plate 8 were made. Dr. E. A. K. Coombs of the University Herbaria, Harvard University, provided valuable bibliographic and technical information.

This thesis was prepared using the TECO and RUNOFF programs running on the DECSystem 1090 installation of the U. W. O. Computing Centre. I am grateful to the Centre's staff for their assistance also with many of the computing aspects of the work reported here.

Finally I would like to thank five people to whom in different ways I am most deeply indebted for my education in all its forms. These are my parents, John Dickinson and Lenore Mikalauskas, and our friends Vera Baker and Paul Williams. My wife, Meher Shaik has helped and encouraged me in every phase of this project. I cannot adequately express here my gratitude for her love and patience.

TABLE OF CONTENTS

(Volume 1)

Page

LIST OF PHOTOGRAPHIC PLATES

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1 "The Crataegus problem" (Palmer, 1932)

This thesis deals with one species group of North American hawthorns, Crataegus crus-galli L. sensu lato, as it occurs in Ontario. This is a group that is readily distinguished from other hawthorns, and one that like the genus as a whole has been split into different numbers of species at various times. The lack of agreement over the past 200 years as to what constitutes a species of Crataegus is the result of interaction between the plant itself, the dialectical development of modern concepts of plant taxa, and the history of North America and its botanical exploration. In the case of this genus in particular, the result is an extreme one (i.e. "The Crataegus Problem" described by Palmer, 1932, 1946, and Camp, 1942, as well as by Rickett, 1936, 1937; Kruschke, 1955, 1965), consisting as described below of too many species distinguished from one another on the basis of criteria and concepts of limited value, and with too little regard for the possible synonymy of many of them with species already described.

The study here of C. crus-galli s. l. in Ontario is an exploration of some of the principal reasons for the taxonomic difficulty encountered with the genus Crataegus in North America. The first of these is the complex morphological variability to be found within the genus. The treatment of Crataegus in Provancher's Flore Canadienne (1862) concluded that "...all of the hawthorns present numerous variations in size, shape and color of their fruit, in the lobing, toothing and shape of their leaves, in the number of styles and nutlets, and in their habit of growth."

Provancher (1862) went on to suggest that this variation, although according to him the result of differences due to age, exposure and habitat, had been responsible for the embarassing degree of synonymy existing then among Crataegus species. Thus, the second source of difficulty is the significance to be attributed to the variation observable in the genus. Slight morphological differences between individuals were uncritically accepted as indicative of differences between species, without sufficient consideration being given to possible alternative explanations. As species of Crataegus continued to be described there was apparently little or no thought given to the nature of the underlying phenomena being described piecemeal in this way.

The third principal source of difficulty today results from the preceding one. Species continued to be described uncritically for several more decades following Provancher's outburst of skepticism, with too little attention paid to establishing their lack of synonymy with species already described (Palmer, 1946). As a result, there are now over a thousand specific names to be considered in any taxonomic revision of the genus in North America.

In the present study only the first two of these three problems will be dealt with, in relation to the results obtained with C. crus-galli s. l. While these results lead to recommendations for future taxonomic treatment, dealing with the problem of synonymy even in a restricted group like C. crus-galli s. l. is beyond the scope of this study, and impracticable anyway since the limited geographic coverage here precludes a taxonomic revision of the entire group at this time.

In the next two sections of this chapter the genus Crataegus L. and C. crus-galli L. are introduced in greater detail. In this account, the taxonomic history of C. crus-galli in North America is taken as far as 1892, the year in which the last conservative treatment of the species, and of the genus as a whole in North America, was published. The events after 1892 are then examined, since they reveal the way in which failure to understand

variability in Crataegus led to nomenclatural chaos. In the final section a plan of the study is given, showing how the questions that it seeks to answer for Ontario C. crus-galli s. l. arise from consideration of the problem described above, of the nature and origin of variability in Crataegus.

1.2 The genus Crataegus L.

The genus Crataegus L., the hawthorn, belongs to the subfamily Maloideae C. Weber. Hawthorns are thorny shrubs or small trees distributed across temperate Eurasia and North America. In eastern North America Crataegus species are probably best known as highly successful weeds of abandoned or poorly managed agricultural land, especially on clay or clay loam soils (Valek, 1980; Phipps & Muniyamma, 1980). Armed with thorns and coppicing readily, hawthorns are very resistant to the effects of browsing, not only by wild animals, but by cattle as well.

Most hawthorns are extremely intolerant of shading, and sites in eastern North America colonized by Crataegus species are probably often occupied for only a single generation. In the absence of repeated disturbance which they are well able to tolerate, hawthorns are shaded out and replaced by other tree genera (Hoover, 1961; Valek, 1980). Prior to Amerindian and European agricultural activity they were probably restricted largely to erosion surfaces,

floodplains and other gaps in the forest (Palmer, 1932, 1946; Marie-Victorin, 1938; Whitney, 1982), habitats which they also continue to occupy (Hoover, 1961; Valek, 1980; Phipps & Muniyamma, 1980).

Recently, El-Gazzar (1980) has divided the genus into two subgenera. One (subgen. Crataegus), comprising the eurasian taxa, is characterized by El-Gazzar as having strongly lobed leaves with "intermediate" veins running from the midrib to the bases of the lateral sinuses. The other (subgen. Americanae El-Gazzar) accounts for the North American taxa and is stated by El-Gazzar to have leaves only shallowly lobed or entire, and lacking intermediate veins. While this classification recognizes a real morphological contrast between much of the Crataegus flora of Eurasia and that of North America, it is based on incomplete and uncritical compilation of the data available (Phipps, in press a, b).

A less simplistic approach to subgeneric classification in Crataegus was that of Loudon (1838), who divided the genus into 15 sections (of which one is now segregated as the genus Pyracantha M. J. Roemer) which he considered to represent natural groups, based on their combined reproductive and vegetative characteristics. Subsequent treatments of the genus, apart from that of El-Gazzar (1980) have for the most part employed a similar approach (Sargent, 1892; Schneider, 1907; Eggleston, 1908a, b; Palmer, 1925,

1963; Kruschke, 1965; Phipps & Muniyamma, 1980).

1.3 Crataegus crus-galli L.

This species is the type of Crataegus section Crus-galli Loudon (= series Crus-galli (Loud.) Rehder). In the widest sense, C. crus-galli L. is now known to be distributed across eastern North America, from southern Ontario and Quebec to northern Florida, west to southern Minnesota, eastern Kansas and eastern Texas (Palmer, 1962, 1963). It was one of the first North American species of the genus to be introduced into European cultivation, at the end of the 17th century (Loudon, 1838). The species was described by Linnaeus from the illustrations and descriptions of Plukenet (1692) and Gronovius (1739).

The Linnaean diagnosis of C. crus-galli appears under Icosandria digynia (Linnaeus, 1753), and is based solely on vegetative characters: "CRATAEGUS foliis lanceolato-ovatis serratis glabris, ramis spinosis." These features are incorporated into Loudon's description of section Crus-galli: "Leaves without lobes, obovate-oblong or obovate-lanceolate, more or less serrated, and of a dark shining green, with petioles margined by the decurrence of the leaf. Fruit small, or middle-sized, round, dark green till ripe, and, when ripe, scarlet. Spines very long, and bent like the spur of a cock."

Loudon grouped into his section Crus-galli three species (C. crus-galli L., C. ovalifolia Horn., and C. prunifolia Bosc.) all described as native to North America and comprising a total of nine varieties all then in cultivation in Britain. These all agreed with the Linnaean species overall (leaves glabrous, unlobed) but at the same time were distinguished from it by details of growth habit, leaf shape, and thorniness. Loudon considered the latter two species to be probably only varieties of C. crus-galli, but his illustrations indicate the presence of correlated variation in floral characteristics such as stamen number. Three of the varieties illustrated (C. crus-galli var. crus-galli and var. salicifolia Aiton, and C. prunifolia Bosc var. prunifolia) all appear to have about ten stamens, while C. crus-galli var. pyracanthifolia Dec. and C. ovalifolia Horn. both have fifteen or more, in the flowers shown.

Crataegus crus-galli L. continued to appear in treatments of the North American flora throughout the 19th century. An extensive synonymy for the species, as well as for five principal varieties recognized earlier on the basis of differences in leaf shape was supplied in Sargent's treatment of Crataegus for volume 4 of his Silva of North America (1892). In addition, C. berberifolia T. & G. was reduced by Sargent to a variety of C. crus-galli, referring

only to the pubescence of its leaves and other parts as a distinguishing feature. Nonetheless, the Linnaean species is illustrated as having about ten stamens per flower (Pl. 178) while var. berberifolia (T. & G.) Sarg. clearly has fifteen or more (Pl. 179).

Loudon (1838) and Sargent, in his first treatment of the genus (1892), both appear to have been inclined to see C. crus-galli L. as a somewhat variable species, in which variation in floral construction occurred but was unworthy of taxonomic recognition. Thus, for most of the 19th century less than a score of Crataegus species and varieties were recognized for North America (Sargent, 1892).

1.4 "The period of expansion" (Palmer, 1932).

The highly conservative treatment of the genus in North America during the nineteenth century was largely the product of the paucity of material available to systematic botanists, even including those based in North America. This situation changed drastically, however, during the last years of that century, with the beginning of what Palmer (1932) later described as "the period of expansion for the genus."

Between 1896 and 1900 three men described a total of 57 species of Crataegus from eastern North America:

W. W. Ashe, 32 species; C. D. Beadle, 17 species; and N. L. Britton, 8 species (Thiselton-Dyer, 1904). In the following ten years Beadle and Ashe alone described a further 260 species (Brown, 1910). Beadle's treatment of the genus for the Flora of the Southeastern United States (1903) recognized 185 species for that area alone. When the situation was reviewed by Brown (1910) a total of 866 new entities had been described since 1896. Beginning in 1901, Sargent alone had published 530 of them (Brown, 1910). Just within section Crus-galli there were by 1925 in excess of a hundred names then current: 71 due to Sargent; 18 to Beadle; and 9 of them to Ashe (Palmer, 1925).

1.5 Charles Sprague Sargent.

In view of the magnitude of his effect on Crataegus taxonomy, it is worthwhile examining C. S. Sargent and his work a little more closely, in order to understand better the situation he helped to create. The following biographical sketch is abstracted from his recent biography by Sutton (1970) and that of Asa Gray, by Dupree (1959). Born in 1841, Sargent was the second son of a wealthy Boston merchant. He attended Harvard College 1858 to 1862, where he was a mediocre student. He apparently did not attend the Natural History lectures given by Asa Gray (Dupree, 1959). Thus the first record of an interest related to plants is

not until after he completed military service during the civil war and then traveled in Europe (1865 - 1868), when he returned to manage his father's estate, Holme Lea, in Brookline, Massachusetts. In becoming interested in landscaping and horticulture at this time Sargent had the approval of family members and friends themselves already committed to such pursuits: not only his father, but also a cousin, Henry W. Sargent, a disciple of the landscape architect A. J. Dowling, as well as his family's friend, H. H. Hunnewell. Sargent himself dated the beginning of his interest in trees to this time when he took charge of Holme Lea.

1.5.1 Creation of the Arnold Arboretum

Meanwhile Sargent's alma mater was seeking a new generation of botanists. Asa Gray had appealed to the president of Harvard in 1871 for relief from his teaching and administrative responsibilities that he had been carrying largely unassisted since 1842. At around the same time, also, Harvard received a bequest from the estate of James Arnold which was to be used for "the promotion of Agricultural, or Horticultural improvements, ..." Gray had been able to have this money earmarked for the creation of an arboretum, and needed "to have somebody got to take charge of tree raising and planting... and to superintend

the whole garden..." (Gray, in Sutton, 1970).

As a result, between 1872 and 1874 four appointments were made: Gray's assistant, the taxonomist S. Watson, became the curator of the herbarium; G. L. Goodale, an M. D. and plant physiologist took over Gray's teaching; and W. G. Farlow, an M. D. as well, and who had studied for two years in Europe with Anton de Bary, became plant pathologist at the Bussey Institute and later Professor of Cryptogamic Botany. The fourth appointment, that of curator of the new arboretum and Professor of Horticulture, was made to Sargent. In 1873 he was made director both of the Arnold Arboretum and of the Botanical Garden in Cambridge. These posts had initially been offered elsewhere, but without being accepted.

According to Gray, the Harvard Corporation had been at a loss to find a suitable candidate (Sutton, 1970). Although no record is available describing how in the end Sargent was selected, the influence of two persons was doubtless of importance (Dupree, 1959; Sutton, 1970). These were his neighbor, the historian Francis Parkman (Sargent's predecessor as Professor of Horticulture, forced by blindness to retire), and H. H. Hunnewell, a friend not only of the Sargents but also of Gray and Eliot (then president of Harvard), as well as a noted Harvard benefactor.

Sargent plunged into his new work with vigor. Having "...brought order out of chaos" at the Botanical Garden (Sargent, quoted by Sutton, 1970), Sargent resigned his directorship there in 1879 in order to concentrate on the arboretum. Meanwhile, he had learned his botany from Asa Gray (Dupree, 1959). He did field work in the western U. S. and prepared a survey of U. S. forests in 1880, as "Expert and Special Agent of the Tenth Census of the United States." In 1883 he also participated in a transcontinental survey of natural resources undertaken by private railroad interests. In 1884 he published a Report on the Forests of North America. He began work on his Silva of North America in 1882, under the sponsorship of the Smithsonian Institution. Volume one appeared in 1891, and the whole ran to a total of fourteen volumes, of which the fourth contained his first treatment of Crategus. During this same period Sargent also directed the collecting for the Jesup Collection of North American Woods for the American Museum of Natural History. In view of this activity, it is perhaps not surprising that when Gray died, Sargent's tribute to his teacher made reference to the way in which Gray "...was too often led away from the main purpose of his life," that is, completion of the Flora of North America (Sutton, 1970). Sargent, after all, was the one who got things done.

1.5.2 Sargent and Crataegus

Sargent's first treatment of Crataegus, in volume four of the Silva has already been mentioned (Section 1.4). The second time that Sargent wrote up Crataegus, for the Supplement to the Silva (1902), the treatment was still not too extreme: a total of only 84 tree or tree-like entities were recognized. However, the following years saw him, and Ashe and Beadle as well, describe more and more new species, as recorded by Brown (1910).

The treatment of the genus by these workers came under attack, however, as other botanists began to question the way in which a genus they had been content to see comprise a score or so of species only a few years before became inflated with hundreds of new names, from North America alone (e. g. Schneider, 1906; references cited by Sutton, 1970).

Sargent replied to these critics in 1907. He based his treatment on the existence of small but distinct discontinuities between entities encountered in the field. He wrote that, "different plants which it had been supposed belonged to one species differed in their time of flowering, in the number of their stamens, in the colour of their anthers, in the time of the ripening of their fruit, and in the nature of the fruit and the form of the nutlets, and

that these characters were constant and could be depended on as distinguishing characters" (Sargent, 1907).

Sargent pointed out that these entities described by him as species "fall into twenty natural groups," groups that corresponded to the sections of the genus described by Loudon plus additional ones created subsequently. Sargent admitted that these groups might by some be considered as the species of Crataegus themselves, the individual entities constituting subspecies and varieties, but apparently felt that his treatment was more convenient. He could offer no explanation of the origin of these distinct entities, however. He ruled out hybridization as a possibility, by reference to the results of seedling trials begun by Beadle, and after 1899, carried out by Sargent as well (see Section 1.6.3, below).

Sargent appears to have taken the position that such cataloguing of nature, without regard to the origin of the diversity encountered, was by itself a worthy task. Its reward lay in the discovery of new plants of economic or aesthetic value, and their introduction into human use. By virtue of belonging to the local elite, an elite largely descended from the original generation of settlers that had remained in place and prospered considerably, Sargent was a political and social conservative (Sutton, 1970). He was undoubtedly influenced by the way in which the evolutionary

biology and genetics of the time were being united with class interests to produce social darwinism and a way of thinking that placed a premium on 'pure' lines, to the denigration of 'mixed' ones (Nordenskiöld, 1928). He was also a scientific conservative, probably because of both his training by Asa Gray and because of his lack of wider academic experience. He remained in apparently complete ignorance of the contemporary European work on apomixis (reviewed by Clausen, 1954) and its potential significance for the interpretation of the results of his seedling trials. Instead, his secure position in society likely contributed to a certain amount of dogmatic self-assuredness on his part, and may have also made it still easier for him to reject the possibility that the species he had discovered were not true species but merely bastards.

1.6 Other Viewpoints.

Brown's 1910 paper had not only summarized the changes that had occurred in hawthorn taxonomy during the preceding fourteen years, but also he had reported the results of a questionnaire that he had circulated among some of the principal students of Crataegus. His questions (Brown, 1910) were as follows:

1. Why did not the systematic botanists discover the large number of species of Crataegus years ago?

2. Do you consider the species now being described elementary species?

3. Do the species breed true or come true to seed?

4. Will different species hybridize?

5. Do you consider the numerous species to have arisen as mutations?

1.6.1 The Failure of the Old Guard

Sargent answered question 1 by saying, "Because they did not use their eyes and were satisfied that what had been published about the genus was correct and final." Brown's other respondents (Ashe, Beadle, W. W. Eggleston, E. Brainerd, and J. Dunbar) proposed more charitable reasons to explain why earlier botanists had failed to recognize so many species. These included lack of adequate herbarium material (Ashe, Eggleston), lack of adequate field study (Beadle), and changing species concepts (Brainerd).

1.6.2 Elementary Species

Brown's second question is a reference to the work of Jordan. Jordan described species of Erophila, Thlaspi, Biscutella, Iberis, and Draba as being made up of many "true" or "elementary" species, very distinct from one

another and highly constant (Brown, 1910; Davis & Heywood, 1965). Sargent replied, "I do not know what you mean by elementary species." Ashe felt that some of the Crataegus species that had been described were elementary species, but that, "When differences [in stamen number or anther color] extend to inflorescence, size of flowers, and foliage, the sum of the correlated characters may be regarded as entitling the form to specific rank." Beadle concurred with Ashe, and pointed also to the duplication which had occurred, through "authors... working altogether too independently of each other's discoveries." Brainerd suggested that "Many or most of the Crataegus entities that had been described were either elementary species or "fluctuations" or "forms."

1.6.3 The Question of Hybridization

All the workers polled referred to the way in which little or no variation was observed among the seedlings of the individuals that served as the basis for description of new species. This was taken as evidence that the species described did indeed breed true. Beadle wrote "...I proved that they came true to seed before daring to publish a new species..," although like all the others except Sargent and Dunbar, he agreed that hybridization probably did occur. With hindsight, Eggleston had the most modern outlook,

having already suggested that "much of the trouble found in Crataegus arises from hybridization, and may not mutation be another disturbing element?" (Eggleston, 1908a). Eggleston's treatment of Crataegus for the seventh edition of Gray's Manual (Eggleston, 1908b) embodied this interpretation by reducing many species to synonymy with others, or to varietal rank (Eggleston, 1908a; Brown, 1910). Ashe allowed that possibly some of the species described represented hybrids. Brainerd referred to the predisposition to hybridization of other Rosaceous genera in suggesting that similar tendencies might be expected in Crataegus, and might yet be found to account for the multitude of stable, elementary species. Brown described the way in which hawthorn colonization of abandoned fields and pastures "...makes cross pollination easier and much more probable" (Brown, 1910). Brown pointed to within-plant variability, the occurrence of individuals intermediate between sympatric species, and the occurrence of very local forms as evidence of hybridization having occurred. However, he admitted that "The fact that Crataegus plants seem to come true to type when grown from seed is a stumbling block in the way of the hybridity theory" (Brown, 1910).

1.7 Recognizing the Significance of Reproductive Behavior

In his 1910 paper Brown had also reported that pollinations between introduced C. monogyna Jacq. and native C. brainerdii Sarg., as well as between a number of other native species all resulted in apparently normal fruit production. Subsequently Standish (1916) demonstrated that pollen fertility varied dramatically among North American Crataegus taxa. Standish concluded that a high degree of pollen sterility was evidence of hybridity, by analogy with other better studied instances, especially among the Rosaceae. Longley (1924) reported on studies of pollen meiosis in which a wide range of North American taxa growing at the Arnold Arboretum were sorted into three groups: diploids, with regular meiosis and normal pollen production; and two types diagnosed as hybrids, each comprising both triploids and tetraploids but differing in the degree of pollen infertility.

Around 1925 E. J. Palmer succeeded Sargent as the principal student of Crataegus in North America. Although he recognized the potential interfertility of Crataegus species, Palmer also pointed out some of the probable barriers to hybridization, and seems not to have been inclined to consider new combinations and synonyms as hybrids (Palmer, 1932, 1946, 1963). In this regard, he also pointed out limitations on the conclusions to be drawn from

the work of Standish and Longley. At the same time, Palmer (1932) was the first to refer to agamospermy as a possible explanation of the constancy observed in seedling lots from individuals otherwise thought to be possible hybrids. He suggested that there was a large group of Crataegus hybrids which had arisen from crosses among triploids and other polyploid forms that were partially or wholly pollen infertile, "but which produce seed apogamously [i.e. apomictically; Rieger, Michaelis & Green, 1968] and therefore reproduce very closely all of the characters of the parent plants..." (Palmer, 1932). Palmer based these conclusions on the results of unpublished experiments performed by Karl Sax at the Arnold Arboretum. These indicated that fruit could be produced in several Crataegus taxa without pollination, by flowers whose stamens and styles had been removed prior to anthesis.

Subsequently, H. W. Rickett (and later Palmer, 1946) made the point that many of the species described earlier by Sargent and others did in fact correspond to forms to be observed in the field. According to Rickett (1936, 1937) the polymorphic species like C. pruinosa and C. crus-galli s. l. that he studied in Missouri were actually complexes of such forms, each one reproducing apomictically. Rickett inferred the occurence of apomixis as did Palmer on the basis of Sax's experimental results, as well as the

occurrence of polyploidy and pollen sterility within these complexes.

Palmer's papers (Palmer, 1932, 1946) and those of Rickett (1936, 1937) had the effect of setting a seal on Crataegus reproductive biology. Although subsequent authors referred to the role of apomixis, hybridization and polyploidy in generating the taxonomic complexity of the genus (Camp, 1942; Camp & Gilly, 1943; Gustafsson, 1947a) there were no further experimental studies of reproduction in North American Crataegus until those reported by Love and Feigen (1978) and here (Chapter 10). Even apomixis itself was not demonstrated anatomically until the work of Muniyamma and Phipps (1979a and unpubl.). Moreover, until the present work all the recent experimental studies of Crataegus reproduction were made on exclusively diploid species, C. monogyna Jacq. and species with which it hybridizes in Britain (C. laevigata (Poir.) DC; Bradshaw, 1953, 1971, 1975) and in Oregon (C. douglasii Lindl. var. suksdorfii Sarg.; Love & Feigen, 1978). Since the hybrids of C. monogyna and other species have proven also to be diploids, where they have been examined (Bradshaw, 1975; Muniyamma & Phipps, 1979b), these studies are unlikely to have provided much information on the role of apomixis in Crataegus hybridization, and in its reproduction in general.

1.8 Plan of the Study

The last part of this introduction has been devoted to reviewing the way in which data concerning the peculiarities of reproductive behavior in Crataegus contributed to a re-evaluation of the Crataegus taxonomy of the 1896-1932 period. Although the role in Crataegus evolution ascribed by earlier workers (e.g. Palmer, 1932, 1946; Rickett, 1936, 1937; Camp, 1942; Camp & Gilly, 1943; Gustafsson, 1947a) to apomixis, hybridization and polyploidy was largely conjectural, much of their surmise has been confirmed by the work of Muniyamma and Phipps (1979a, b, and unpubl.), as well as that described here in Chapters 9 and 10. However, apart from the work of Rickett (1936, 1937) and more recently that of Sinnott, on C. section Pruinosae in Ontario (Sinnott, 1978; Sinnott & Phipps, 1983), little attention has been paid to detailed documentation of the morphological variation occurring in North American Crataegus.

Resolution of the taxonomic aspects of "The Crataegus Problem" (Palmer, 1932; Camp, 1942) also requires an understanding of the nature and origin of the patterns of variability found in the genus. Accordingly, while documenting reproductive behavior in Ontario C. crus-galli s. l., this thesis has concentrated on an exposition of the patterns of morphological variation and variability in this

group, and of the methods suitable for their study.

The plan of the study is illustrated in Figure 1.1, where it can be seen to consist of four parts: choice of taxa and sites, description of variation, description of reproductive phenomena, and finally, a synthesis of the results obtained.

1.8.1 Sampling Considerations

The Crataegus crus-galli L. sensu lato complex of taxa was selected for study for the following reasons: (1) the group exhibits the high degree of the morphological differentiation for which the genus has become notorious; (2) it has been shown to be a polyploid complex in Ontario (Muniyamma & Phipps, 1979b); (3) it is highly distinctive, and not readily confused with any other group; and (4) in southern Ontario it is both common and locally abundant.

The importance of the first reason is self-evident. The second reason independently suggests that the group exhibits processes such as apomixis previously ascribed to the genus but until recently examined neither anatomically nor experimentally. The last two reasons given are pragmatic ones. Since intensive study of several local samples is required in order to obtain detailed data on both morphological variation and reproductive behavior, the group

chosen should not require excessive time to be spent in locating and sampling stands. The criteria used in selecting sites were similarly pragmatic, and are discussed in Chapter 2.

Characterization of the phenetic variability and reproductive behavior of C. crus-galli s. l. as described here requires some kind of comparative basis. This has been provided by a parallel though less extensive examination of C. punctata Jacq. For purposes of comparison with C. crus-galli s. l. the important feature of this species is the way in which it is much less sharply differentiated internally, both cytologically and morphologically. Crataegus punctata appears to be uniformly diploid where it has been studied (Muniyamma & Phipps, 1979b), and shows little of the discontinuous variation (e. g. in stamen number and other correlated characteristics) that is so conspicuous in C. crus-galli s. l. The distribution of C. punctata is broadly comparable to that of C. crus-galli s. l., although it does not extend as far south, and extends farther north (Palmer, 1963; Phipps & Muniyamma, 1980).

1.8.2 Description of Phenotypic and Nucleotypic Variation

Phenotypic variation in C. crus-galli s. l. and C. punctata was documented from voucher collections made from the individual trees sampled at each site. As

described in Chapter 3, this involved first selection and scoring of descriptors, followed by preliminary analyses to describe and summarize the data so obtained.

Subsequent analyses embodied two aspects of the model for the analysis of the structure of taxonomic collections described by Orlóci (1968), namely cluster-seeking and searching for trends in variation. Such an approach, employing techniques for the analysis of multivariate data, was felt to be best suited to describing the complex morphological variation encountered in Crataegus species groups like C. crus-galli s. l. or C. section Pruinosae (Sinnott & Phipps, 1983).

The group structure of the sample was examined by means of cluster analyses (Chapters 5 and 6) based on matrices of resemblance coefficients for multivariate data described in Chapter 4. Trends among groups with respect to variation in the descriptors used were examined by ordination techniques described in Chapter 6. The covariation of these descriptors was analyzed as described in Chapter 7 in order to determine whether the degree of variability or the pattern of covariation differed among the groups found in the sample. Finally, in Chapter 8 ordination and other methods are used to statistically evaluate the distinctness of these groups.

Study of nucleotypic variation in C. crus-galli s. l. (Chapter 9) was confined to determining chromosome numbers for a subsample of individuals stratified by site and taxonomic affiliation.

1.8.3 Description of Reproductive Phenomena

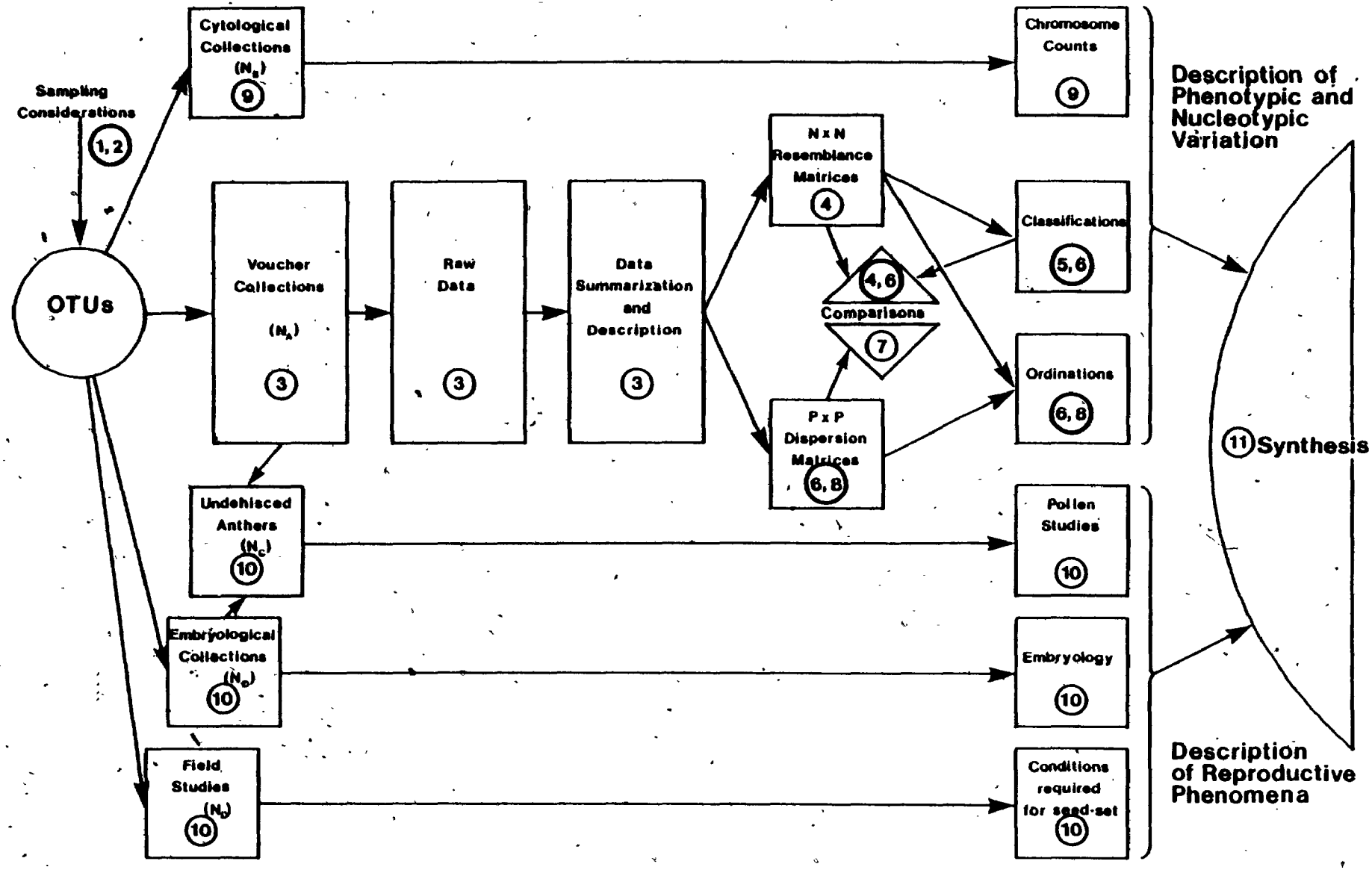
In examining reproductive phenomena in C. crus-galli s. l. (Chapter 10) attention was focused on documenting (a) pollen fertility and production; (b) the occurrence of apomixis; and (c) the conditions under which seed-set occurred in the field. Combination of the results of the descriptive (a, b) and experimental (c) examination of reproduction in C. crus-galli s. l. and C. punctata made it possible to distinguish which of the alternative breeding systems (predominant outcrossing, predominant selfing, and apomixis with or without pseudogamy) was prevalent in each of the taxa studied.

1.8.4 Synthesis

Synthesis of the results of examining variation on the one hand, and reproductive behavior on the other leads in Chapter 11 to adapting to C. crus-galli s. l. a model originally proposed by Camp and Gilly (1943) to describe evolution in polyploid Crataegus species groups in general. This is followed by a discussion of the conclusions to be

drawn concerning the study of *Crataegus* in the past and in the future, in light of the proposed model. Finally, recommendations are made concerning the taxonomic treatment of *C. crus-galli s. l.*, on the basis of the results obtained and the model they support.

Figure 1.1 Plan of the study presented in this thesis. As described in Chapter 2 the OTUs are individual trees of *Crataegus* taxa. Numbers in circles associated with each element of the study indicate the chapters describing that element. N_A, N_B, N_C and N_D are the sample sizes associated with the five parts of the study ($N_A > N_C > N_B$ and N_D); these are explained in the relevant chapters.


Sampling Considerations
1, 2
OTUs
Cytological Collections (N_B)
9
Voucher Collections (N_A)
3
Raw Data
3
Data Summarization and Description
3
N x N Resemblance Matrices
4
4, 6
Comparisons
7
P x P Dispersion Matrices
6, 8
Undehisced Anthers (N_C)
10
Embryological Collections (N_D)
10
Field Studies (N_D)
10
Chromosome Counts
9
Classifications
5, 6
Ordinations
6, 8
Pollen Studies
10
Embryology
10
Conditions required for seed-set
10
Description of Phenotypic and Nucleotypic Variation
11 Synthesis
Description of Reproductive Phenomena

CHAPTER TWO

SITE SELECTION AND SAMPLING METHODS; SITE DESCRIPTIONS

2.1 Introduction

In the present study the sampling unit was an individual hawthorn tree. Such sampling units are the operational taxonomic units (OTUs) with which the study deals, and will usually be referred to in this way, or as 'individuals'.

The principal objective of the procedures described in this chapter was to provide samples from local aggregations of Crataegus crus-galli s. l. and C. punctata individuals for intensive study of the morphological variation and variability present in these aggregations. Unbiased (random) sampling is required in order to justify making the inference that the samples obtained are representative of the aggregations from which they were drawn.

At the same time studies such as the present one may also be used to refine reference classifications available. This requires extensive rather than intensive sampling in order to ensure covering as much of the range, and as many major variants as possible. For these purposes sampling can

be deterministic.

Sampling sites themselves (Fig. 2.1) were selected deterministically. Most of the sites were selected because they were known from the field work of J. B. Phipps to have an abundance of Crataegus crus-galli s. l. or C. punctata. Most of those studied intensively (Section 2.5) are close to London, Ontario and are located in such a way as to minimize the temporal overlap of the very short flowering periods at each site, given the typical progression of spring in the area (Webber & Hoffmann, 1967). Site 4 was chosen because of the occurrence there of not only a distinctive form of C. crus-galli s. l., but also of a putative crus-galli x punctata hybrid (Phipps & Muniyamma, 1980) Other sites sampled deterministically (Section 2.6) were selected because they represented important elements in the Ontario distribution of section Crus-galli (Sites 7,8), or because they permitted inclusion in the sample of individuals of interest, such as ones whose chromosome number was already established (Sites 10, 12, Muniyamma & Phipps, 1979b). Site 6 was studied because of the presence there of another unique form belonging to section Crus-galli. Site 13 was included because of the occurrence there of the sole extant individual (Stewart 2506) identified by Phipps and Muniyamma (1980) as C. disperma Ashe, another putative crus-galli x punctata hybrid (Palmer,

1963). The sampling sites used in this study thus cover virtually the entire Ontario range of Crataegus section Crus-galli (compare Fig. 2.1 with Maps 2, 3 and 4 in Phipps & Muniyamma, 1980).

Topographic maps used in this study are from the National Topographic Series, published by the Department of Mines and Technical Surveys, Ottawa. Air photos were obtained from the Department of Energy, Mines, and Resources, Ontario Ministry of Natural Resources. Data on soil type at each site are taken from county soil maps prepared by the Experimental Farms Service, Canada Department of Agriculture, Ottawa.

2.2 Terminology

Each of the following two sections explains the use in this thesis of a term that may be unfamiliar to some readers, or one whose usage here is one to which they may be unaccustomed.

2.2.1 Morphotype

The term 'morphotype' has been used in this thesis to describe groups of OTUs which all share features of a broadly interpreted species, Crataegus crus-galli L. sensu lato, but which differ from each other in a number of

conspicuous correlated morphological and phenological characteristics. Use of such an informal subspecific category is a convenience because it recognizes the groupings of OTUs that are obvious, *a priori*, but does so without committing the user to a particular nomenclatural solution. However, nomenclatural recommendations have been made in Chapter 11, in order to integrate the results obtained with existing treatments of the genus, notably the recent one for Ontario by Phipps and Muniyamma (1980).

The groups of OTUs recognized *a priori* within *C. crus-galli s. l.* in this study and referred to here as morphotypes exhibit two kinds of geographic distributions. Two are each presently known only from a single locality. One of these is the putative hybrid found at Site 4 that has been referred to *C. ?grandis* Ashe by Phipps and Muniyamma (1980). The other, from Site 6, has been referred by them to *C.* sp. aff. *C. bushii* Sarg. Both morphotypes are referred to by these names in the remainder of the study.

The remaining two morphotypes depart much less than the preceding ones from an idea of *C. crus-galli s. str.*, and have a much wider distribution in southern Ontario. They are distinguished from each other most obviously by the number of stamens per flower and their relative time of anthesis. The first of these is referred to as the '10-stamen' morphotype of *C. crus-galli s. l.* and

characteristically has 5 - 15 stamens per flower. The second is referred to as the '20-stamen' morphotype, has characteristically 10 - 20 stamens per flower, and flowers earlier than the 10-stamen one, where the two occur together.

2.2.2 Topodeme

The term 'topodeme' (Gilmour & Gregor, 1939) has been used in this thesis to refer to local aggregations of OTUs considered to be taxonomically homogeneous at the level of species or morphotype. As originally defined, the -deme terminology was meant to describe "any assemblage of taxonomically closely related individuals" (Gilmour & Gregor, 1939). 'Topodeme' has been employed here in order to explicitly avoid implying anything about the relationships among the OTUs concerned except their common location and taxonomic homogeneity. Since one of the questions addressed by this thesis is that of the nature and consequences of the reproductive behavior prevailing in assemblages of hawthorns it is important to avoid the use of collective terms whose usage may be such that they may connote particular reproductive or other relationships among the included individuals. A similar study of the genus Potentilla L. in Britain (Smith, 1963a, b, 1971) has used the term in the same way.

The word 'population,' a possible alternative to topodeme, was rejected because it is often used in such a way as to imply the occurrence of interbreeding among component individuals. Thus in their revision of the -deme terminology Gilmour and Heslop-Harrison (1954) referred to the possibility of "...looking upon the whole population as the ultimate gamodeme..." Van Valen (1976), in describing an ecological species concept, defines a population as a group of individuals among which interbreeding occurs.

Gilmour and Heslop-Harrison redefined the root 'deme' in part "...so as to indicate even more unequivocally that no idea of 'population' enters into the definition..." (Gilmour & Heslop-Harrison, 1954). For this reason the terminology they propose and 'topodeme' in particular is especially suited for the present study since it enables convenient reference to groups of OTUs of the same morphotype from the same site, without implying anything about their genetic or reproductive relationships. The term is particularly convenient to use in southern Ontario because of the way in which hawthorn stands (Crataegus topodemes) tend often to be spatially well-defined as a result of extrinsic features of land ownership and use. For the purposes of this thesis topodemes are further defined as the aggregation of individuals of a particular taxon at one of the study sites described below (Section 2.5, 2.6).

Topodeme samples are collections of component individuals drawn from within a defined area at certain of these sites, according to the methods described below (Section 2.3).

2.3 Random Sampling

At Sites 1 through 5 (Section 2.5) random samples of OTUs were drawn by mapping the site from air photos or topographic maps and establishing a sampling frame; a grid was superimposed on the frame and twenty to forty sampling points were selected using a random number table (Rohlf & Sokal, 1969) to supply grid coordinates. Each sampling point was located on the ground, and the nearest Crataegus (crus-galli or punctata, depending on the site) over 2m high was marked permanently and sampled. The only exception to this method was Site 1, with the smallest area, where a complete mapping and enumeration of all Crataegus over 2m in height was possible in the spring of 1978. This enabled drawing a random sample directly from the enumeration. Although the sites varied in size, the attempt was made to obtain from each site twenty to forty OTUs, without regard to morphotype (in the case of C. crus-galli s. l.). It proved impossible, however, to obtain complete data sets from all of the OTUs sampled at any given site because of the vagaries of flowering, or the disease or death of OTUs. As a result, sampling intensity at these sites varied from

4.0 OTUs per ha (Site 3) to 18 OTUs per ha (Site 1). This variation was a function also of the varying density of the hawthorn cover at the sites, that at Site 3 being the most sparse (Plate 1).

Randomly sampled OTUs were given unique numbers indicating the site at which they are located: Site 1, OTUs 101-134; Site 2, OTUs 201-234 (first sampling) and 401-430 (second sampling); Site 3, OTUs 301-316; Site 4, OTUs 601-622; and Site 5, OTUs 501-520. In the descriptions of the randomly sampled sites which follow (Section 2.5) the topodeme samples (identified as T1 - T7) drawn at each one are indicated. Individual OTUs are identified according to taxon in the plans of Sites 1 - 5 (Fig. 2.2 - 2.6) and according to both site and taxon in Table 5.1 (Chapter 5). Appendix 1 lists the voucher specimens deposited in the Herbarium of the Department of Plant Sciences, University of Western Ontario (UWO) for all the randomly sampled OTUs studied here.

2.4 Deterministic sampling

At the other sites (6 - 13; Section 2.6), and to a limited extent at Sites 2 - 5 as well, small numbers of additional OTUs were chosen deterministically. As outlined above, the reason for choosing some of these was the availability of published information about them, either

cytological or taxonomic. Additional OTUs were selected at Site 2 because of the availability of data concerning them from preliminary studies. Three additional OTUs of C. punctata were selected, at Sites 2 and 3. The random sample at Site 4 included two individuals of C. ?grandis. The remaining individuals of this morphotype were enumerated and incorporated into the total sample, as were two individuals intermediate between C. ?grandis and the distinctive form of 10-stamen C. crus-galli also present at this site. At these and other sites it was a matter also of obtaining additional OTUs to represent variants believed to exist but which had not appeared in the random samples, or of selecting representative OTUs at sites with biogeographic or other signficance. Especially in the case of the sites more distant from London (Sites 6 - 8), the availability of flowers at the time of collection was of paramount importance in selecting OTUs.

Deterministically sampled OTUs are given unique numbers indicating the site at which they are located: at Site 4, additional OTUs are numbered 691-697; and at Site 6, the sample is numbered 801-815. The remaining deterministically sampled OTUs are numbered 701-793, using the second digit to indicate the particular site in question. The individual OTUs present at sites 6 - 13 are indicated in the descriptions of these sites (Section 2.6), as well as in

Table 5.1 (Chapter 5). Appendix 1 contains a list of the voucher specimens deposited in UWO for all the deterministically sampled OTUs studied here.

Figure 2.1 Map of southern Ontario showing locations of the randomly sampled topodemes (T1 - T7) at Sites 1 through 5, as well as those of the additional Sites 6 through 13. See Sections 2.5 and 2.6, respectively, for complete descriptions of these sites. Legend of the map shows the symbols used to represent the topodeme samples here and in the illustrations of the following chapters.

CRATAEGUS CRUS-GALLI STUDY SITES IN SOUTHERN ONTARIO

RANDOMLY SAMPLED TOPODEMES

- □ T1 SITE 1 *Crataegus crus-galli* (20-stamen)
- ■ T2 SITE 1 *Crataegus crus-galli* (10-stamen)
- ○ T4 SITE 2 *Crataegus crus-galli* (20-stamen)
- ● T5 SITE 2 *Crataegus crus-galli* (10-stamen)
- ▲ T3 SITE 3 *Crataegus crus-galli* (10-stamen)
- ★ T6 SITE 4 *Crataegus crus-galli* (10-stamen)
- P T7 SITE 5 *Crataegus punctata*

Species and morphotypes represented by additional, deterministically sampled individuals:

Crataegus crus-galli (10-stamen) SITES 6, 7, 8, 9, 10, 12

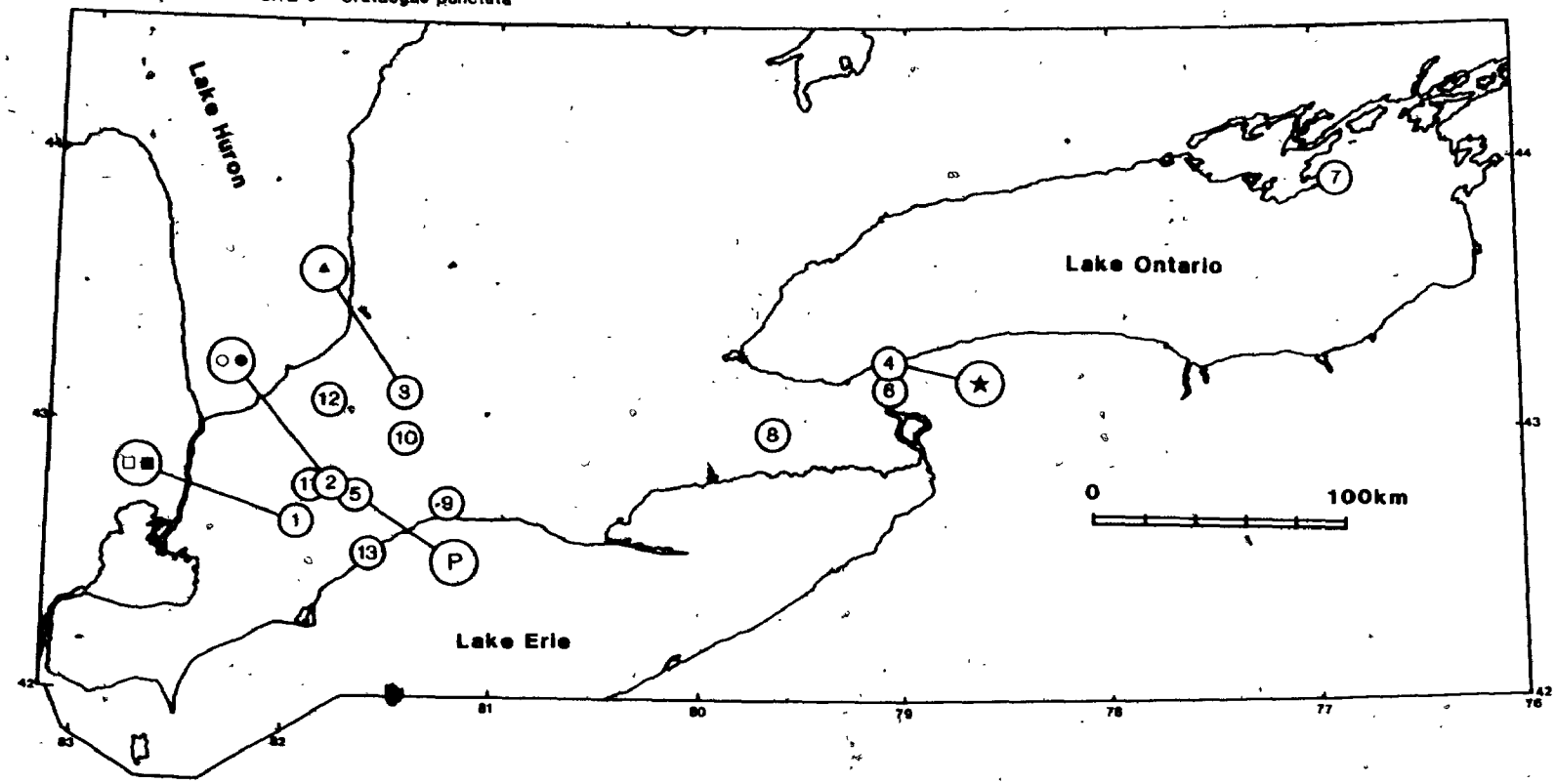
Crataegus crus-galli (20-stamen) SITES 9, 11

Crataegus sp. aff. *C. bushii* SITE 6

Crataegus ?*grandis* SITE 4.

Crataegus punctata SITES 2, 3.

Crataegus Series Macracanthae SITES 5, 13.



2.5 Randomly sampled sites

2.5.1 Site 1, Fansher Road (Figure 2.1, 2.2; Plate 1c)

Lambton Co., Euphemia Twp. Conc. 5.

42.66° N Lat., 81.94° W Long.

Topographic map reference (1:25000) 40I/12E.

Military Grid reference 233235.

Elevation 640 feet above mean sea level.

Soil type: Brady sand.

Sampling frame 1.5 ha.

Topodeme Samples T1 and T2.

This site is owned by Mr. C. Gawne of Florence, Ontario, and is used by him as rough pasture for cattle. The site runs E-W, bounded on the south by Fansher Creek and on the north by a cultivated field (Fig. 2.2; Plate 1c). Much of the site is open, with clumps of hawthorns on the higher ground, and grass on the slopes and in the swales. Willows form thickets along the creek. At the west end of the site an elm thicket has developed and has shaded out a number of the hawthorns originally tagged there in 1978. Several Juniperus virginiana L. are present along the northern boundary of the site, especially at the eastern end. These probably account for the high frequency of infestation of hawthorns here by Cedar Apple Rust (Appendix 5), particularly on the earlier flowering 20-stamen

C. crus-galli.

The site was originally found by Phipps, who made a number of collections there. Two of the C. crus-galli s. l. documented cytologically by Muniyamma and Phipps (1979b) are located here: Phipps 4454 and 4719, both determined to be triploids.

The sampling frame was established in 1978 and within it all Crataegus 2m high or over were mapped and tagged. A total of 301 trees were recorded in this way. Using a random number table (Rohlf & Sokal, 1969) a sample of 40 OTUs was drawn in 1978. In 1979, in order to find additional individuals for experiments on reproductive behavior (Section 10.5), a further 22 individuals were drawn from the enumeration. Of the 62 randomly chosen OTUs, 53 were Crataegus crus-galli sensu lato. Of these, 33 were the 10-stamen morphotype and 17 were the 20-stamen morphotype, while 3 could not be classified because they failed to flower. The remaining 9 individuals in the sample were C. punctata Jacq. (1), C. compacta Sarg. (3), C. schuettei Ashe var. basilica (Beadle) Phipps (1), and C. calpodendron (Ehrh.) Medic. (2).

Availability of complete collections led to the use of only 27 individuals from this sample, 17 10-stamen OTUs and 10 20-stamen ones (including Phipps 4719).

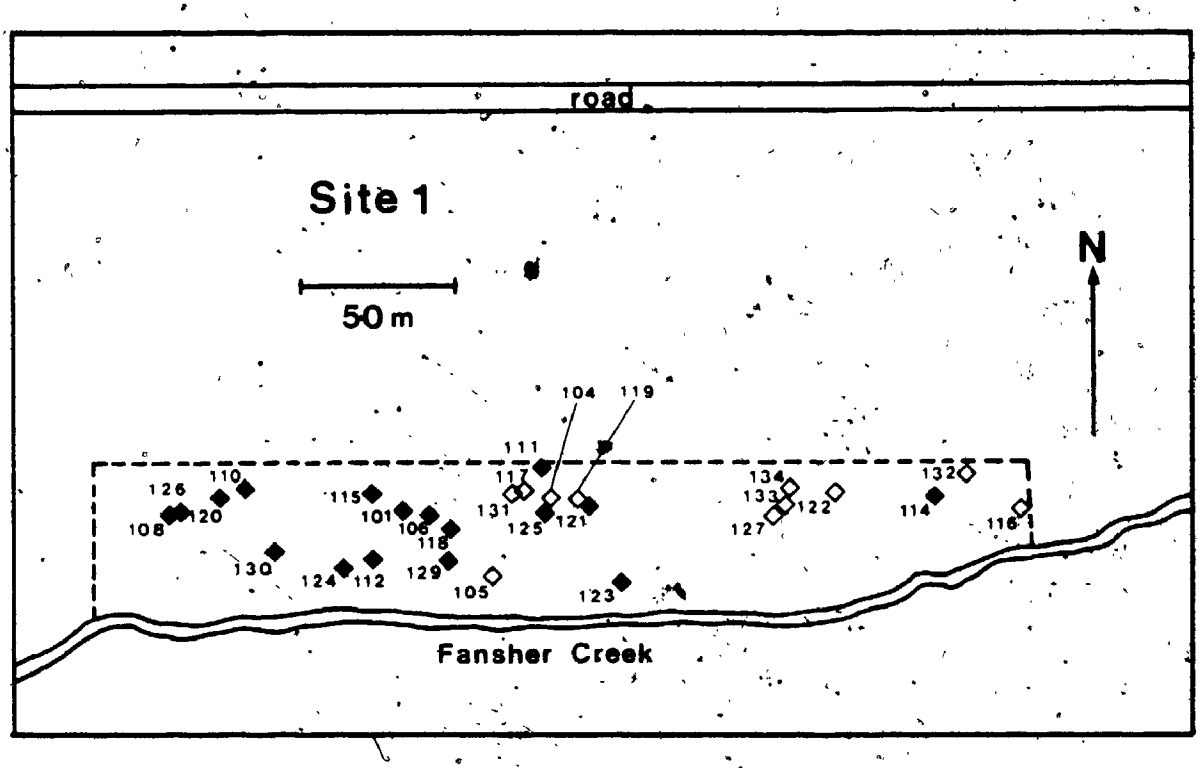


Figure 2.2 Site 1, Fansher Road. Dashed line and creek delimit sampling frame. Diamonds indicate locations of randomly sampled OTUs. Open symbols, topodeme sample T1 (20-stamen C. crus-galli); solid symbols, topodeme sample T2 (10-stamen C. crus-galli). Compare Plate 1c.

2.5.2 Site 2, Mosa Township (Fig. 2.1, 2.3; Plate 1e).

Middlesex Co., Mosa Twp. Conc. 8.

42.77° N Lat., 81.81° W Long.

Topographic map reference (1:25000) 40I/13c.

Military Grid reference 333355.

Elevation 700 feet about mean sea level.

Soil type: Brookston clay loam.

Sampling frame 4.7 ha.

Topodeme Samples T4 and T5.

This site is owned by Mr. A. Carrothers of Appin, Ontario, who rents it out for rough pasture. It is a half concession width lot extending about 500m along Mosa Concession 7-8 road, and consists of three main areas: bottom land along a creek flowing northwest into the nearby Sydenham River; a small area of high ground southwest of this creek; and a larger area of high ground along the northeast side of the creek. Hawthorn thickets now cover both areas of high ground, and there are scattered hawthorns on the more open bottom land. In a few places forest trees (maple, beech) remain, principally at the tops of the slopes. The land has belonged to Carrothers' family since the latter part of the 19th century, and was farmed into the first part of the 20th century. Following abandonment of cultivation, hawthorns invaded but were then cleared out completely except for a few individuals left as shade trees

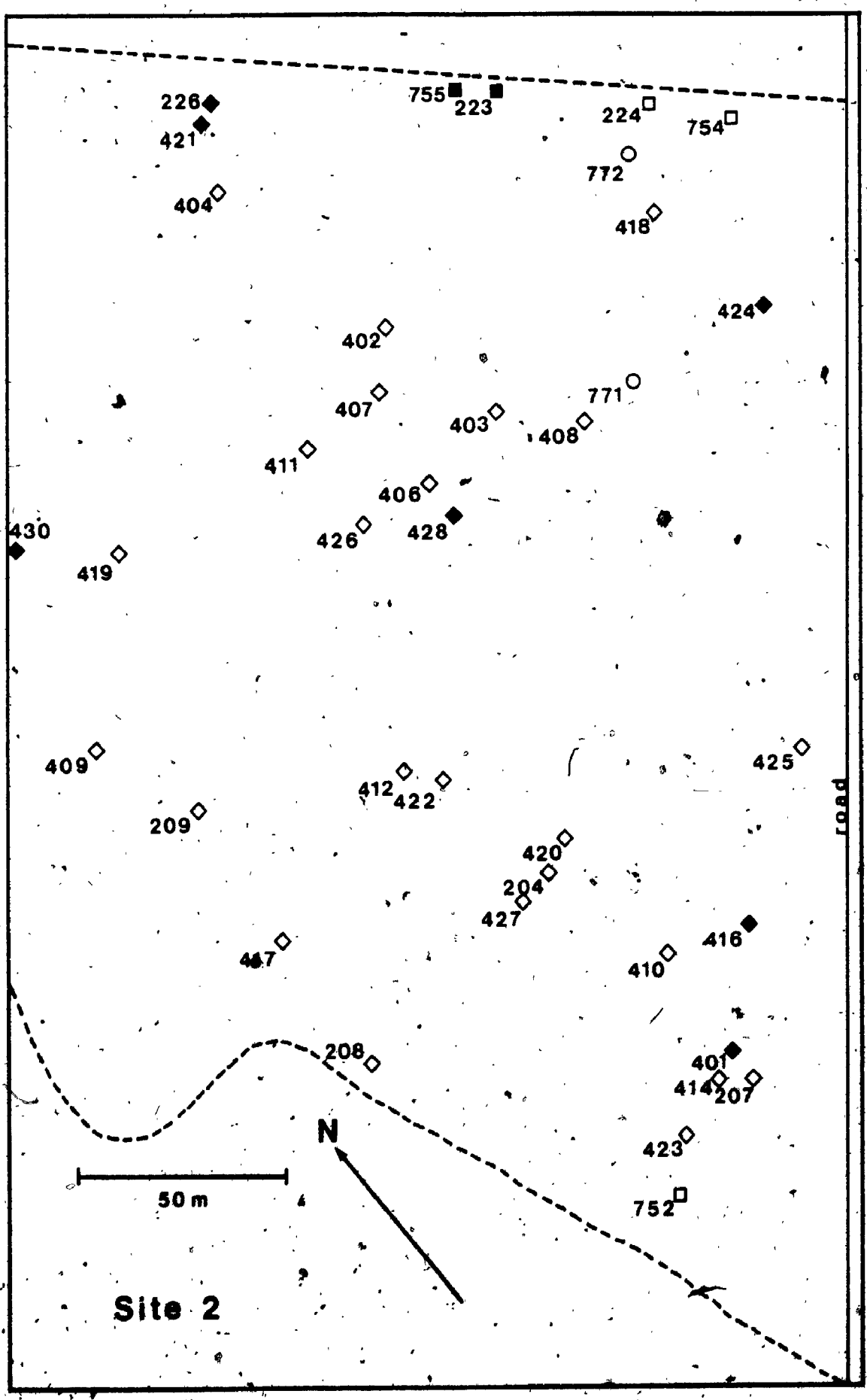
for cattle. An individual of C. punctata on the NW boundary of the property has a diameter at breast height of approximately 40-45 cm, and probably represents one such pasture shade tree. These trees are alleged to have been responsible for the reinvasion of hawthorns that resulted in the present thickets. A second attempt was made to remove hawthorns from the site, probably in the late 1940's. This time only a strip about 200 m wide was cleared, along the concession road. Even this area now supports a well established hawthorn cover, although differences in structure between it and the less recently disturbed area can be detected (Plate 1e).

The site is particularly rich in 20-stamen C. crus-galli. Several trees in the eastern corner of the property were vouchered for Phipps in 1977 by D. Barrows (see Appendix 1). The larger area of high ground northeast of the creek was selected for sampling. In 1978 a 22.75 ha sample area was surveyed, extending across the entire lot (Plate 1e). Trees of C. crus-galli s. l. were located at 28 randomly chosen grid intersections. Of these, 9 were the 10-stamen morphotype, and 13 were the 20-stamen one. Six trees could not be determined to morphotype because flowering material could not be collected.

In order to increase the sampling intensity within at least a portion of this site, a part of the 1978 sampling

area was resurveyed in 1979 (Fig. 2.3; Plate 1e). In this area, which corresponds approximately to the portion of the site cleared most recently, an additional 30 OTUs were located in the same way as before. This area also contained 7 OTUs from the first sample. Of this sample of 37 OTUs, 26 are 20-stamen, 9 are 10-stamen C. crus-galli and two could not be determined to morphotype. Of these, 7 10-stamen and 24 20-stamen OTUs make up the two random samples from this site for which complete data were obtained. A number of deterministically selected C. crus-galli, some of them originally vouchered by Barrows and three C. punctata were also added to the sample here (Table 5.1; Appendix 1).

Figure 2.3 Site 2, Mosa Township. Dashed line, road and frame of the illustration (at left) delimit the sampling frame. Diamonds indicate locations of randomly sampled OTUs of 10-stamen (solid symbols; topodeme sample T5) and 20-stamen (open symbols; topodeme sample T4) C. crus-galli. Deterministically sampled OTUs are indicated by squares (C. crus-galli s. l.; morphotypes coded as with T4 and T5) and circles (C. punctata). Compare Plate 1e.


226
421
755
223
224
754
772
404
418
424
402
771
407
403
408
411
406
430
426
428
419
409
425
412
422
209
420
204
427
416
410
208
401
414
207
423
752
N
50 m
Site 2
road

2.5.3 Site 3, Denfield Side Road (Fig. 2.1, 2.4; Plate 1a)

Middlesex Co., London Twp. Conc. 16.

43.14° N Lat., 81.43° W Long.

Topographic map reference (1:25000) 40P/3e.

Military Grid reference 646757.

Elevation 890 feet above mean sea level.

Soil type: Huron clay loam.

Sampling frame 4.0 ha.

Topodeme Sample T3.

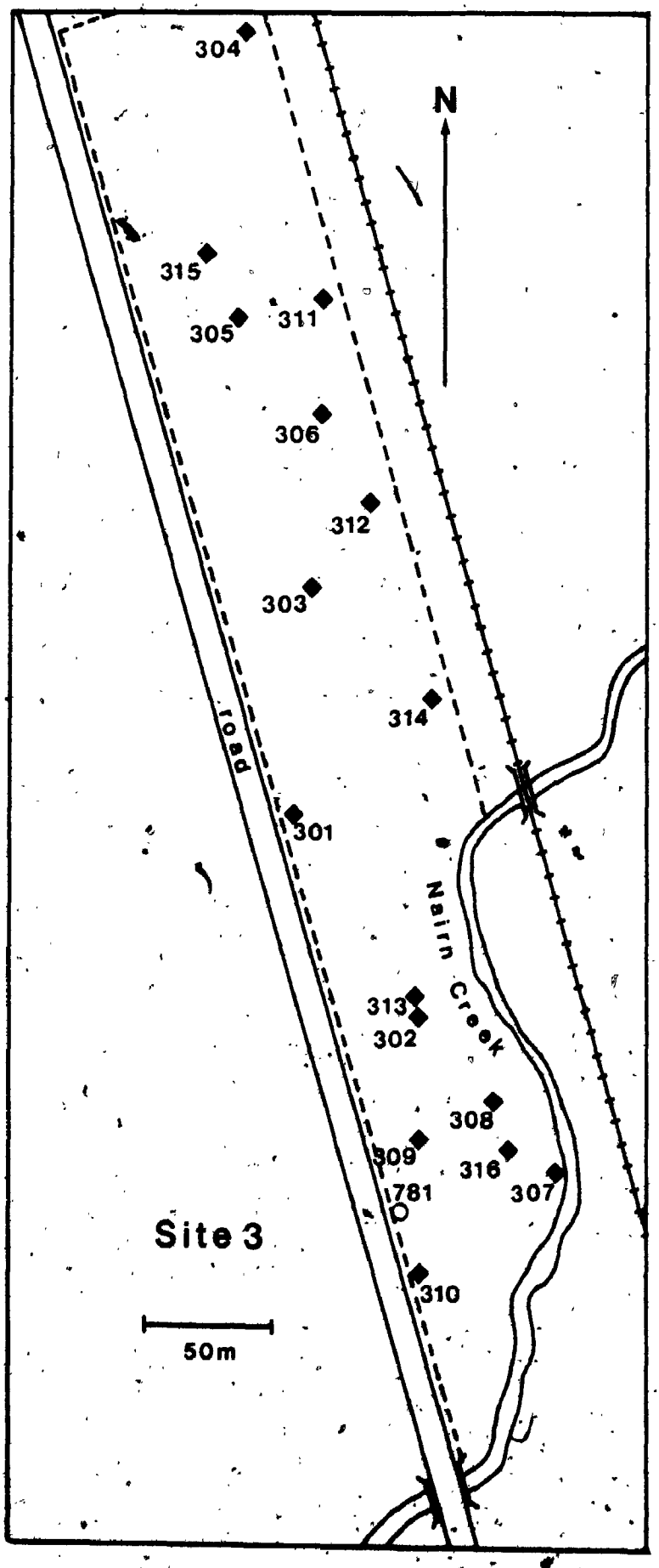
This site was owned by the Ontario Pork Producers Marketing Board when it was surveyed in 1978. In 1980 or 1981 title apparently passed to the present owners, New Life Mills. The site is a narrow strip of abandoned farmland lying between Denfield Side Road and the Canadian National Railway right-of-way south of Highway 7 and immediately to the north of Nairn Creek (Fig. 2.4; Plate 1a). In 1981 construction of a feed mill began, which destroyed several OTUs. In the spring of 1982, the entire site was cleared of hawthorns.

The site is the one most sparsely colonized by hawthorns (Plate 1a). A likely seed source for this and other nearby colonizations is an area of mature hawthorn thickets immediately to the east, across Nairn Creek (Plate 1a). Species diversity is low, with trees in the area being mostly 10-stamen C. crus-galli, and with only scattered

individuals of C. punctata and C. monogyna Jacq.

Since only 10-stamen C. crus-galli was present at this site a sample of only 20 OTUs was intended. However, construction of a tile drain across the north end (Plate 1a) shortly after the initial survey curtailed this plan, and in the end only 16 OTUs were completely vouchered, in the undisturbed area of the site.

Figure 2.4 Site 3, Denfield Side Road. Dashed line and creek delimit sampling frame. Diamonds indicate locations of randomly sampled OTUs of 10-stamen C. crus-galli (topodeme sample T3). Deterministically sampled OTU of C. punctata indicated by circle. Compare Plate 1a.


304
N
315
311
305
306
312
303
314
road
301
Nairn Creek
313
302
308
309
316
307
781
Site 3
310
50m

2.5.4 Site 4, Fort George National Historic Park (Fig. 2.1, 2.5; Plate 1b).

Regional Municipality of Niagara, Niagara-on-the-Lake.

43.24° N Lat., 79.06° W Long.

Topographic map reference (1:25000) 30M/3h.

Military Grid reference 575898.

Elevation 300 feet above mean sea level.

Soil type: Haldimand clay loam.

Sample frame 3.6 ha.

Topodeme Sample T6.

This site is on the grounds of Fort George National Historical Park at Niagara-on-the-Lake. It is situated for the most part between a grove of oaks ("Pardise Grove") and an abandoned railroad right-of-way. The area in which the site is located has been a military reservation and public common since the 18th century, and a national park since 1937. Consequently, it is quite possible that hawthorns have been a part of the local vegetation since shortly after the land was first cleared. Hawthorns are present now almost exclusively as large individuals with a diameter at breast height usually 20 - 30 cm, 3 - 5m high, and with a very broad crown, where growing in the open. Along the railroad right-of-way (presently a large ditch) hawthorn cover is much more dense, forming a closed canopy along the eastern bank. The hawthorns present are principally a

distinctive form of 10-stamen C. crus-galli with red anthers, identified by Sargent (1908) from collections made by J. Dunbar as C. crus-galli var. pyracanthifolia Ait. In addition there are also present seven or eight individuals of the crus-galli morphotype identified by Phipps and Muniyamma (1980) as C. ?grandis Ashe, as well as several individuals of C. conspecta Sarg. Establishment of new individuals at this site appears to be restricted because of periodic mowing by Parks Canada personnel.

At this site, because of its irregular shape, a modified sampling method was used. Within the sampling frame (Fig. 2.5) a grid of possible sampling points was laid down, 8m apart, in a systematic sequence. Starting with the first point, a random number of points were counted off in order, and the next point was marked for sampling. This was repeated until a total of 32 possible sampling points had been chosen. These were located on the ground, and the nearest Crataegus section Crus-galli was found. If there was no such tree within 15m the point was declared empty. In this way a total of 22 OTUs were located. Twenty of these were 10-stamen C. crus-galli, of which 19 yielded complete data. Two of the individuals of C. ?grandis at this site also were in the random sample. The other individuals of this morphotype present here were also vouchered, as were two individuals intermediate between

C. ?grandis and 10-stamen C. crus-galli (OTUs 696 and 697). In addition, two C. conspecta were also vouchered as they were used as pollen sources in experiments described in Section 10.5.

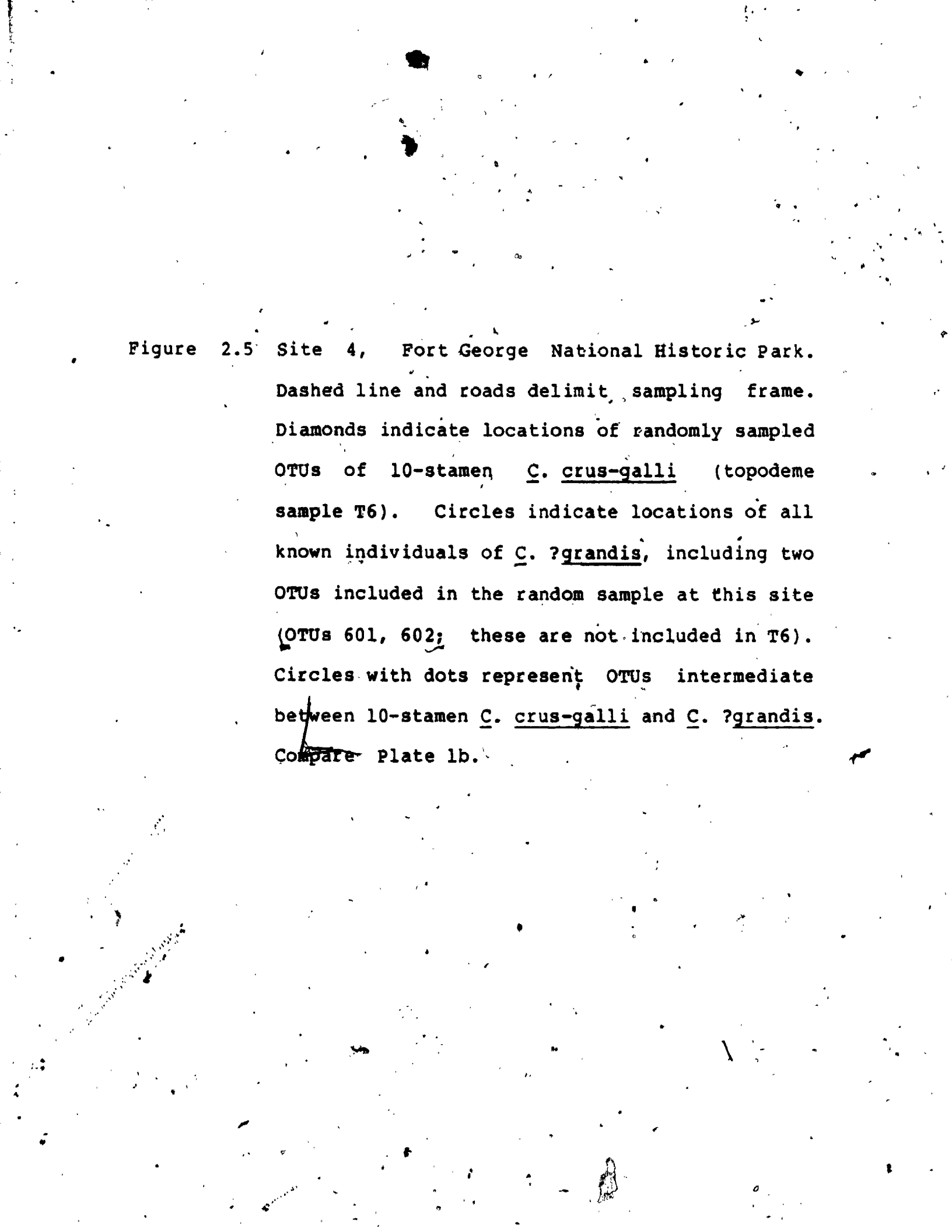
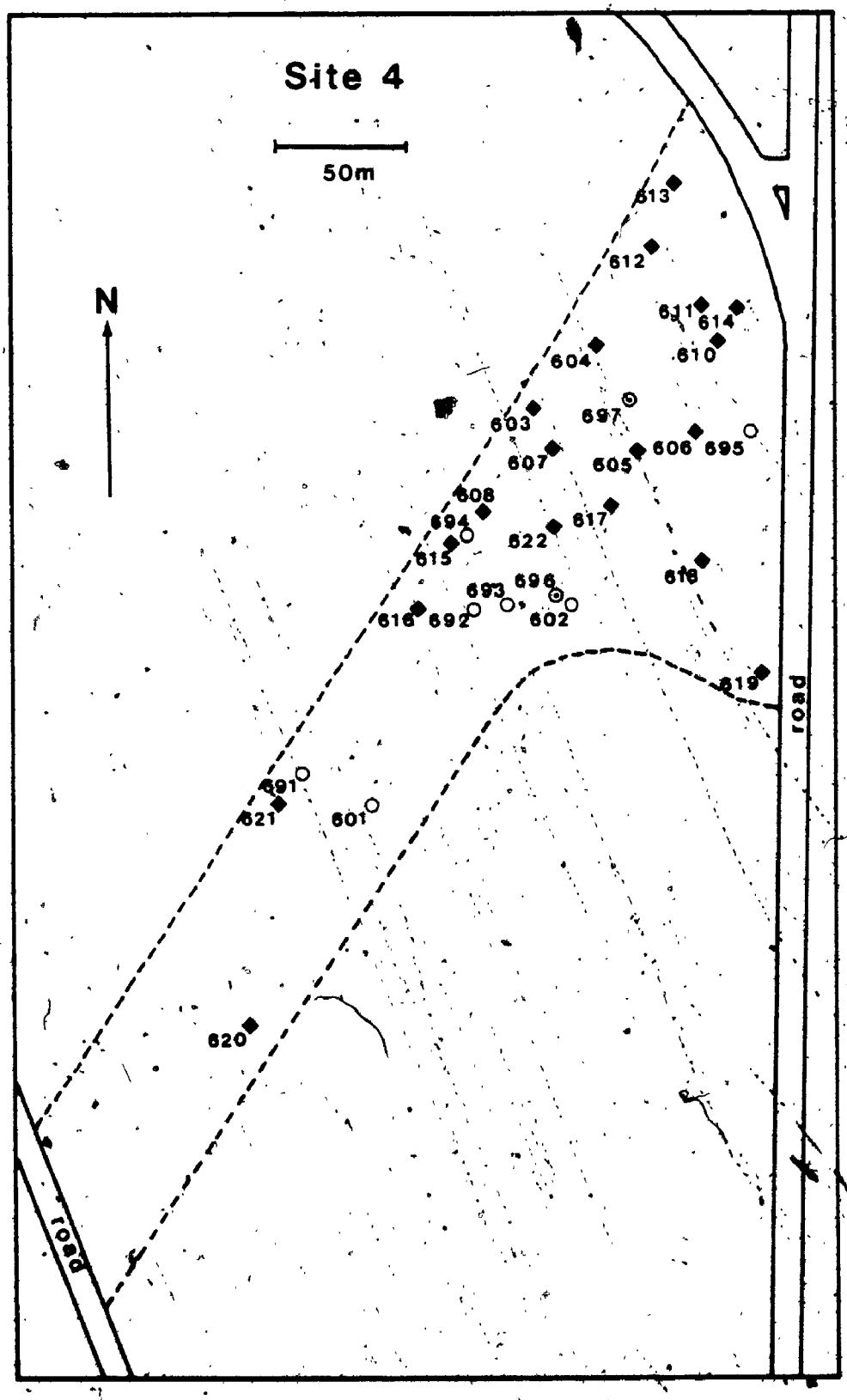


Figure 2.5 Site 4, Fort George National Historic Park. Dashed line and roads delimit sampling frame. Diamonds indicate locations of randomly sampled OTUs of 10-stamen C. crus-galli (topodeme sample T6). Circles indicate locations of all known individuals of C. ?grandis, including two OTUs included in the random sample at this site (OTUs 601, 602; these are not included in T6). Circles with dots represent OTUs intermediate between 10-stamen C. crus-galli and C. ?grandis. Compare Plate 1b.


Site 4
50m
N
613
612
611
614
610
604
603
697
607
606
695
605
608
694
617
622
615
618
693
696
616
692
602
619
road
691
621
601
620
road

2.5.5 Site 5, Newbiggen Creek (Fig. 2.1, 2.6; Plate 1d)

Middlesex Co., Ekfrid Twp. Range 1 N.

42.73° N Lat., 81.67° W Long.

Topographic map reference (1:25000) 40I/12g.

Military Grid reference 453306.

Elevation 680 feet above mean sea level.

Soil type: Haldimand clay loam (and bottom land).

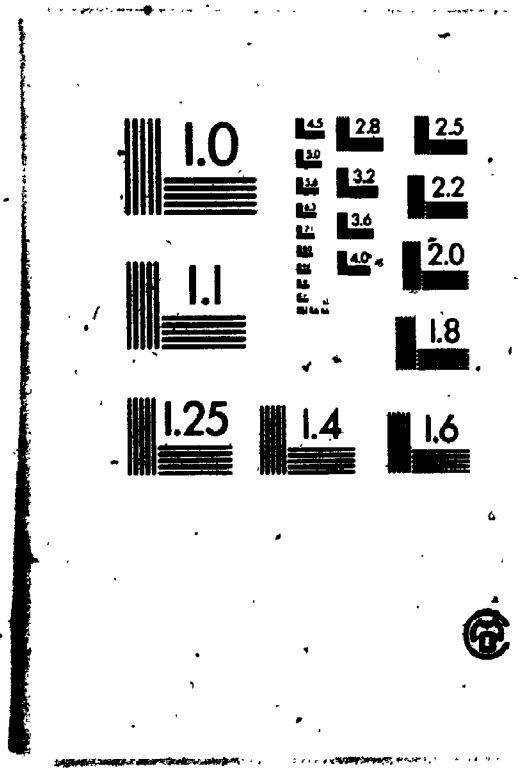
Sampling frame 2.7 na.

Topodeme Sample T7.

Ownership of this site was not determined. The site is located on a bend in Newbiggen Creek on the north side of Highway 2, southeast of Glencoe. The presence of a slough separating the two areas of hawthorn cover (Plate 1d) suggests that the eastern area may at one time have been on the opposite side of an oxbow from the western area. In the spring this slough is regularly filled with standing water, which may account for the fact that no successful hawthorn colonization is evident in it.

This site was sampled for C. punctata, although this species is not extremely abundant here. Most of the Crataegus cover is in fact due to C. crus-galli s. l., together with C. mollis (T. & G.) Scheele and other species. These latter include C. succulenta Link (OTU 769).

2



As at Site 3, only a single taxon was sampled at this site, and so only 20 sampling points were located. Of these, complete data were obtained for the trees at only 18.

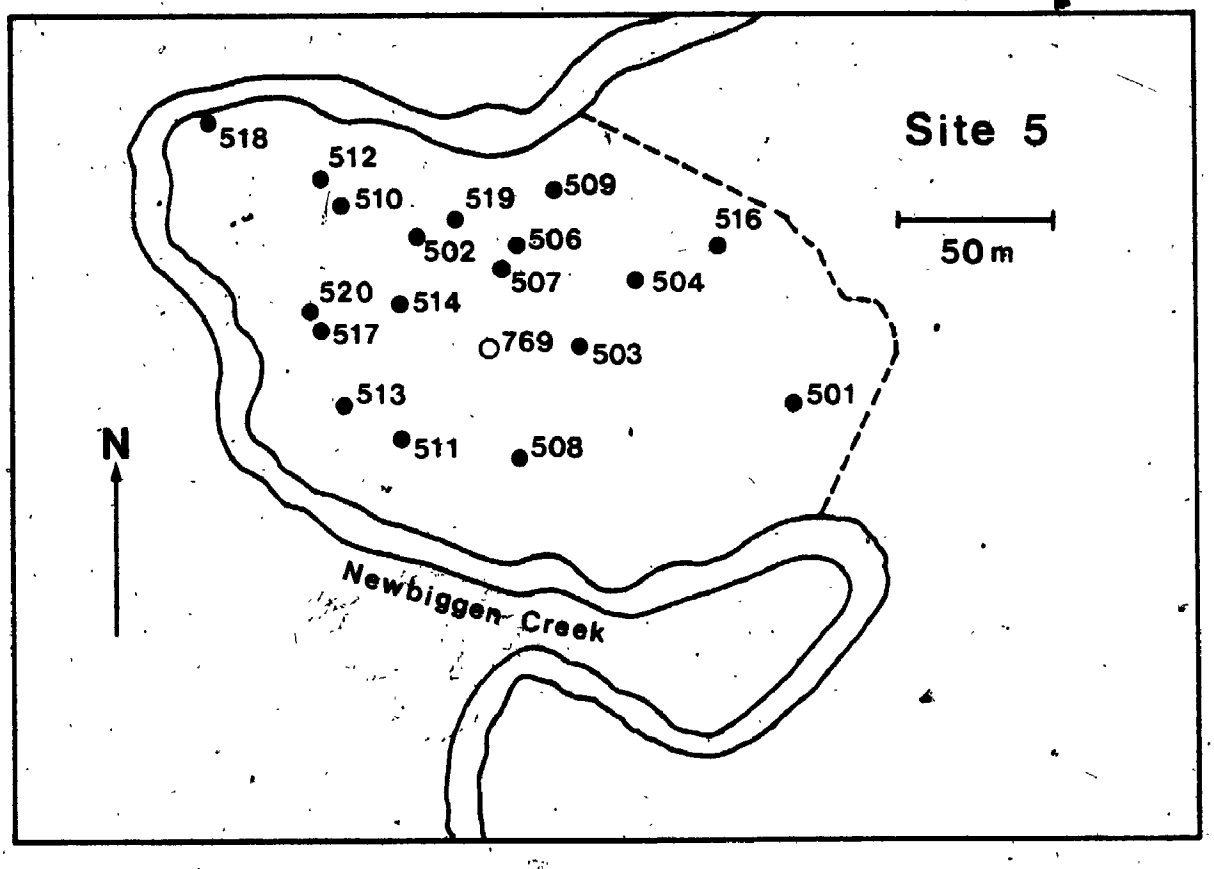
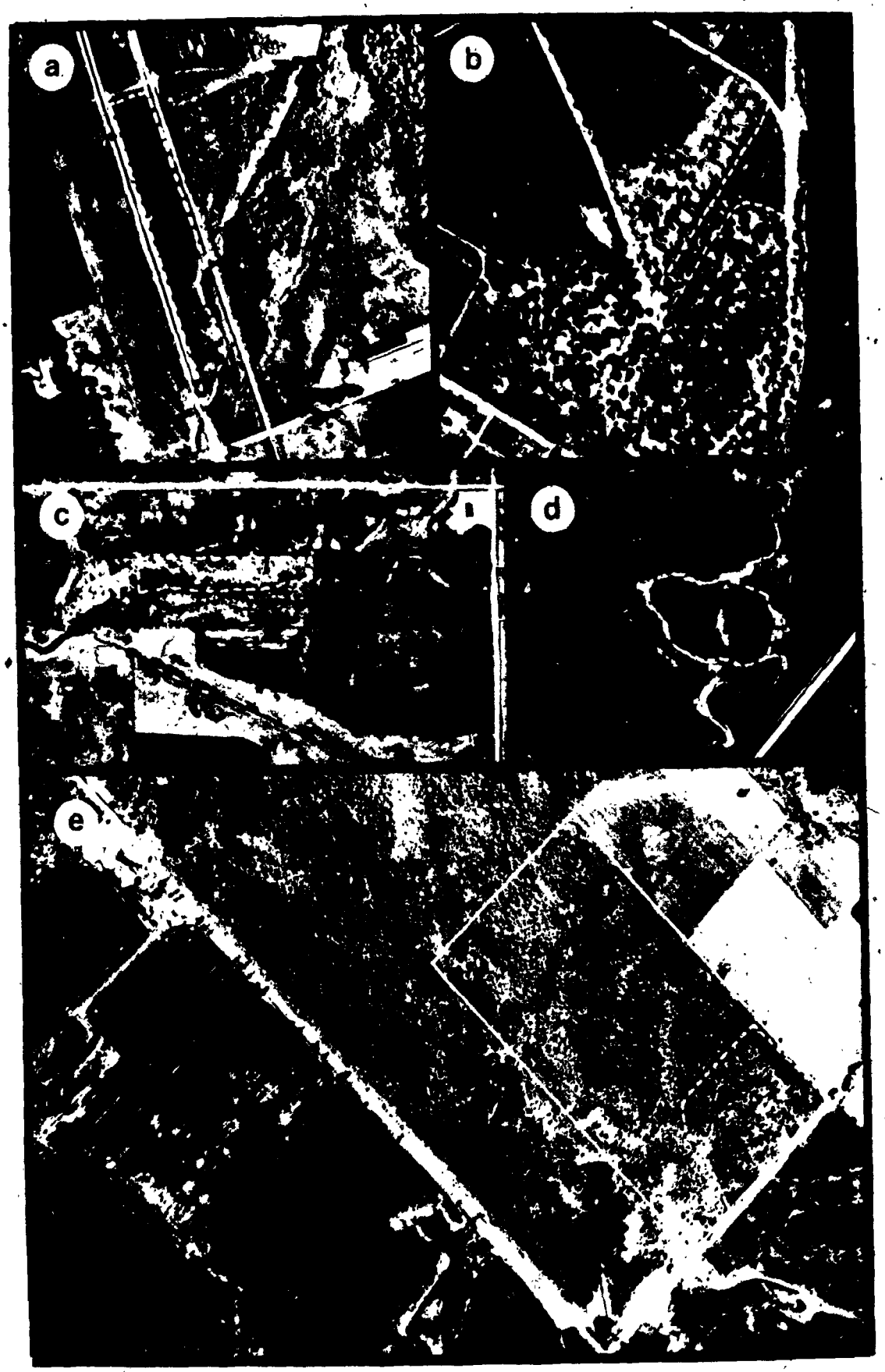


Figure 2.6 Site 5, Newbiggen Creek. Dashed line and creek delimit sampling frame. Solid circles indicate locations of randomly sampled OTUs of C. punctata (topodeme sample T7). Open circle indicates location of deterministically sampled individual of C. succulenta. Compare Plate 1d.

Plate 1 Aerial views of the randomly sampled sites. Source: Ontario Ministry of Natural Resources. Scale approximately 250m to 1 cm. In (a - e) north is at the top of the page. Sampling frames indicated by dashed lines. (a) Site 3, Denfield Side Road (OMNR 1978, flight line 4310, roll 233, negative 57). Compare Figure 2.4. (b) Site 4, Fort George National Historical Park (OMNR 1978, flight line 4317, roll 29, negative 490). Compare Figure 2.5. (c) Site 1, Fansher Road (OMNR 1978, flight line 4246, roll 254, negative 76). Compare Figure 2.2. (d) Site 5, Newbiggen Creek (OMNR 1978, flight line 4251, roll 203, negative 75). Compare Figure 2.6. (e) Site 2, Mosa Township (OMNR 1978, flight line 4254, roll 227, negative 2). In (e), solid line outlines the area originally surveyed. Compare Figure 2.3.


a
b
c
d
e

2.6 Deterministically Sampled Sites

2.6.1 Site 6, Queenston (Fig. 2.1)

43.16° N Lat., 79.07° W Long.

Topographic map reference (1:25000) 30M/3h.

Military Grid reference 573808.

Elevation 350 feet above mean sea level.

Soil type: Lincoln clay.

OTUs vouchered: 801 - 815.

This site, immediately to the west of the Niagara Parkway in Queenston, is comprised of abandoned agricultural land (orchard, fields) and adjacent fencerows, all colonized to a greater or lesser extent by hawthorns. The taxa most common are 10-stamen C. crus-galli and the crus-galli morphotype identified by Phipps and Muniyamma (1980) as C. sp. aff. C. bushii.

2.6.2 Site 7, Prince Edward Point (Fig. 2.1)

Prince Edward Co., South Marysburgh Twp.

43.93° N Lat., 76.87° W Long.

Topographic map reference (1:25000) 30N/15f.

Military Grid reference 502660.

Elevation 255 feet above mean sea level.

Soil type: Farmington loam.

OTUs vouchered: 791 - 793.

This site is located in part within the Prince Edward Point National Wildlife Area administered by the Canadian Wildlife Service (Environment Canada). This and a nearby site outside Milford, Prince Edward County, appear to be the northernmost Ontario stations of C. crus-galli s. l..

2.6.3 Site 8, Oswego Creek (Fig. 2.1)

Regional Municipality of Haldimand-Norfolk,

Canboro Twp. Conc. 1.

42.99° N Lat., 79.67° W Long.

Topographic map reference (1:25000) 30L/13g.

Military Grid reference 083605.

Elevation 575 feet above mean sea level.

Soil type: Haldimand clay.

OTUs vouchered: 702 - 704.

This site is a rough pasture with hawthorn thickets, on the north bank of Oswego Creek, just east of Canboro.

2.6.4 Site 9, Kettle Creek (Fig. 2.1)

Elgin Co., Yarmouth Twp. Conc. 6.

42.74° N Lat., 81.20° W Long.

Topographic map reference (1:25000) 40N/11g.

Military Grid reference 833324.

Elevation 650 feet above mean sea level.

Soil type: bottom land.

OTUs vouchered: 711 - 714.

This site is a large hawthorn thicket on a knoll above Kettle Creek. It is presently being studied by Mr. P. G. Smith on account of the abundance of Crataegus section Rotundifolieae, but it also contains a number of individuals of both 10- and 20-stamen C. crus-galli.

2.6.5 Site 10, Komoka (Fig. 2.1)

Middlesex Co., Lobo Twp. Conc. 2.

42.96° N Lat., 81.41° W Long.

Topographic map reference (1:25000) 40I/14e.

Military Grid reference 667562.

Elevation 820 feet above mean sea level.

Soil type: Burford gravelly loam.

OTUs vouchered: 731 (= Phipps 4598).

This site is a small area of mature and younger hawthorns in a cattle pasture on the north side of County Road 14, just east of its intersection with County Road 38. It contains several trees vouchered by Phipps, including one determined by Muniyamma and Phipps (1979b) to be a diploid (Phipps 4598).

2.6.6 Site 11, Fencerow near Mosa (Fig. 2.1)

Middlesex Co., Mosa Twp. Conc. 8.

42.76° N Lat., 81.83° W Long.

Topographic map reference (1:25000) 40I/13c.

Military Grid reference 325346.

Elevation 700 feet above mean sea level.

Soil type: Brookston clay loam sand spot phase.

OTUs vouchered: 751 (= Barrows 17).

This fencerow on the north side of Mosa Concession 7-8 Road contains individuals of both common morphotypes of C. crus-galli, including individuals cited as exemplar specimens of C. fontanesiana (Spach) Steud. (Barrows 17) and C. sp. aff. C. livoniana (Phipps 4824) by Phipps and Muniyamma (1980).

2.6.7 Site 12, Hungry Hollow (Fig. 2.1)

Middlesex Co., West Williams Twp.

43.08° N Lat., 81.80° W Long.

Topographic map reference (1:50000) 40P/4.

Military Grid reference 350691.

Elevation 660 feet above mean sea level.

Soil type: bottom land.

OTUs vouchered: 721 (= Phipps 4604, 4637) and 722 (Phipps 4607a).

This site consists of hawthorn thickets on waste land ("Hungry Hollow") beside the Ausable River, outside Arkona. It contains many trees vouchered by Sinnott (Crataegus section Pruinosae) and by Phipps, including two 10-stamen C. crus-galli individuals (Phipps 4604, 4607a) determined to be a tetraploid and a triploid, respectively (Muniyamma & Phipps, 1979b).

2.6.8 Site 13, Port Glasgow (Fig. 2.1)

Elgin Co., Aldborough Twp.

42.51° N Lat., 81.61° W Long.

Topographic map reference (1:25000) 40I/12a.

Military Grid reference 494060.

Elevation 610 feet above mean sea level.

Soil type: Berrien loamy sand.

OTUs vouchered: 761 (= Stewart 2506) and 762.

This site is a fencerow beside a hillside pasture above the Beattie Access Area at Port Glasgow. Hawthorns present include an individual cited by Phipps and Muniyamma (1980) as an exemplar of C. disperma Ashe (Stewart 2506).

CHAPTER THREE

DATA COLLECTION AND SUMMARIZATION

3.1 Introduction

This chapter describes first the kind of collections that were made for all OTUs in the total sample. Next, the morphometric data compiled from these collections are discussed: the descriptors selected, and the way in which they were scored. Finally, the last section of the chapter describes the data obtained, giving descriptive statistics for each topodeme random sample as well as for the samples of C. ?grandis and C. sp. aff. C. bushii. Details of the way in which some of these collections were used in exclusively embryological studies are deferred until Chapter 10. Similarly, the use of collections made entirely for cytological purposes is described in Chapter 9.

With one exception, the descriptors chosen to be the basis for comparisons between OTUs and between topodeme samples are ones which have already been found to be useful in Crataegus taxonomy (Kruschke, 1955; Phipps & Muniyamma, 1980). These are leaf size and shape; characteristics of leaf secondary venation; leaf and calyx lobe margination; flower size; number of androecial and gynoecial parts;

color of undehisced anthers; and fruit size and shape. More emphasis is placed on measurements of selected dimensions of the structures studied than appears to have been the case in the past, however. The only new descriptor found is the presence or absence of a projection of the stamen filament between the anther lobes (comparable to the "distal appendage" illustrated by Hufford, 1980, for Asarum canadense L.).

The two approaches taken to summarizing the compiled data are described in Section 3.6. For the first it was assumed that the scores for the individual descriptors were distributed more or less normally, and that hence the OTU-descriptor mean provided an adequate summary of the data by means of which OTUs could be compared. The second was to use the observed distribution of the data as a basis for comparison between OTUs, employing an information theoretical measure (Section 4.4).

3.2 Data Collection

Each OTU in the total sample of 160 Crataegus used in this study is represented by a pair of herbarium vouchers, one collected in flower, in May or June, and the other in fruit, in September or October. In addition, for each OTU a sample in almost all cases of a minimum of 10 unopened and 10 open flowers was collected at the time of anthesis and

stored in 70% ethanol (see Chapters 9 and 10 for details of the fixatives used). Likewise, a minimum of 20 (usually 50-100) mature fruit was collected for every OTU. In collecting both flowers and fruit the attempt was made to include material from as much of the surface of the canopy as possible. Data on the color of undehisced anthers at the time of anthesis (ANTH; Table 3.1) was collected in the field, without replication.

A midsummer collection of all the leaves from 6-10 individual short shoots (either reproductive or sterile) was also made for each OTU, being careful to record the sequential position along the shoot of each leaf. Here too the attempt was made to select only material from the relatively unshaded surface of the canopy.

With all collections, great care was taken to ensure that the material collected really belonged to the intended OTU. This was especially important where OTUs grew in dense fencerows or clumps, with branches from a number of individuals criss-crossing in a common crown. In some cases in order to avoid confusion material was collected only from the side of an OTU away from adjacent individuals.

3.3 Inflorescence and Flower Descriptors

A total of 9 descriptors of the inflorescence axes and individual flowers was scored (Table 3.1). These can be divided into three categories according to data type: meristic (STYL, STAM); binary (PROJ) and multistate (ANTH, TCAL, PUB1, PUB2); and continuous (WFL and LCAL). The states recognized for the binary and multistate descriptors, and the corresponding numerical scores, are indicated in Table 3.1.

Inflorescence axis pubescence (PUB2) was scored without replication from the spring herbarium vouchers. Style number (STYL), stamen number (STAM), calyx lobe toothing (TCAL), and ovary pubescence (PUB1) were scored on a total of 20 unopened and open flowers (Fig. 3.1). The presence or absence of an apical projection of the stamen filament (PROJ) was scored on the undehisced anthers of 10 unopened flowers only. The dimensions WFL and LCAL were measured to the nearest 0.5 mm on 10 open flowers for each OTU (Fig. 3.1).

In order to maximize topodeme sample sizes when data collection ceased in 1981, only those OTUs for which either flower or fruit collections were lacking were excluded from the study. This resulted in missing data for PUB2 in three out of 111 OTUs, and for ANTH in ten out of 111 OTUs. None

of the deterministically sampled OTUs included in the study had any data missing. In the case of the randomly sampled OTUs, topodeme sample means were calculated from the data available and substituted for missing value codes in the 13 instances where they occurred. This approach has been used in other studies (e.g. Phipps, 1970) and is recommended by Green (1979), on the basis of results obtained by Chan (1972). Since neither of these descriptors were highly variable within topodeme samples, and since both have subsequently been excluded from the analyses concerned with testing hypotheses about group structure (Chapter 7, 8), any adverse effect of this replacement of missing data is likely to be outweighed by the resulting benefits of the larger sample size.

3.4 Fruit Descriptors

Prior to being scored, fruit was stored in plastic bags containing damp paper towelling (samples from individual OTUs contained in separate paper bags), at approximately 5° C. Under these conditions (Kruschke, 1955) fruit of both C. crus-galli s. l. and C. punctata remained firm and unshrivelled for several months. Fruit were processed for scoring by bisecting them longitudinally, inking the cut surface of the pericarp on a stamp pad, and printing it on a 5 x 7 inch card. Two such cards accommodated prints of both

halves of 20 fruit, together with the corresponding pyrenes, extracted, cleaned, and glued down. Thus a permanent record was prepared of fruit size and shape, as well as pyrene number. The cards were photographed so as to provide a record of pyrene appearance, prior to destructive analysis. Seed extraction was done by cutting into the abaxial surface of the pyrene longitudinally with a jeweler's saw, and then using secateurs to split the pyrene along its adaxial suture and the cut. Presence or absence of seed was also discovered by cutting the pryene in half transversely with the secateurs, and examining the contents of the lumen. Criteria for scoring seed present or absent are presented in Section 10.5.2.

Fruit dimensions LFR and WFR (Table 3.1, Fig. 3.1) were recorded to the nearest 0.5 mm from the fruit prints. Pyrene number was also scored. Pyrene number was not used as a descriptor since it is very highly correlated with STYL, i.e. with the number of gynoecial units per flower. Pyrene appearance was virtually uniform throughout both C. crus-galli s. l. and C. punctata, except as it was affected by the variation in the number of pyrenes per fruit. Only in one topodeme sample (T6) did pyrenes - like the fruit themselves - appear to be characteristically more elongate than in other samples. Pyrenes are useful, however, in distinguishing sections Crus-galli or Punctatae

from sections such as the Macracanthae, in which the pyrenes are deeply excavated on their adaxial surfaces.

Figure 3.1 Illustration of selected flower descriptors: (a - d) states of calyx lobe margination, TCAL; (a) 0, (b) 1, (c) 2, and (d) 3. Flower (e) and fruit (f) dimensions as indicated: WFL, flower width; LCAL, length of longest calyx lobe; LFR, fruit length along longitudinal axis; WFR, fruit width perpendicular to LFR. Abbreviations in (e) and (f): G, style; K, calyx lobe; p, pyrene; and xA, position of removed stamen. See Table 3.1 for complete explanation of descriptors.

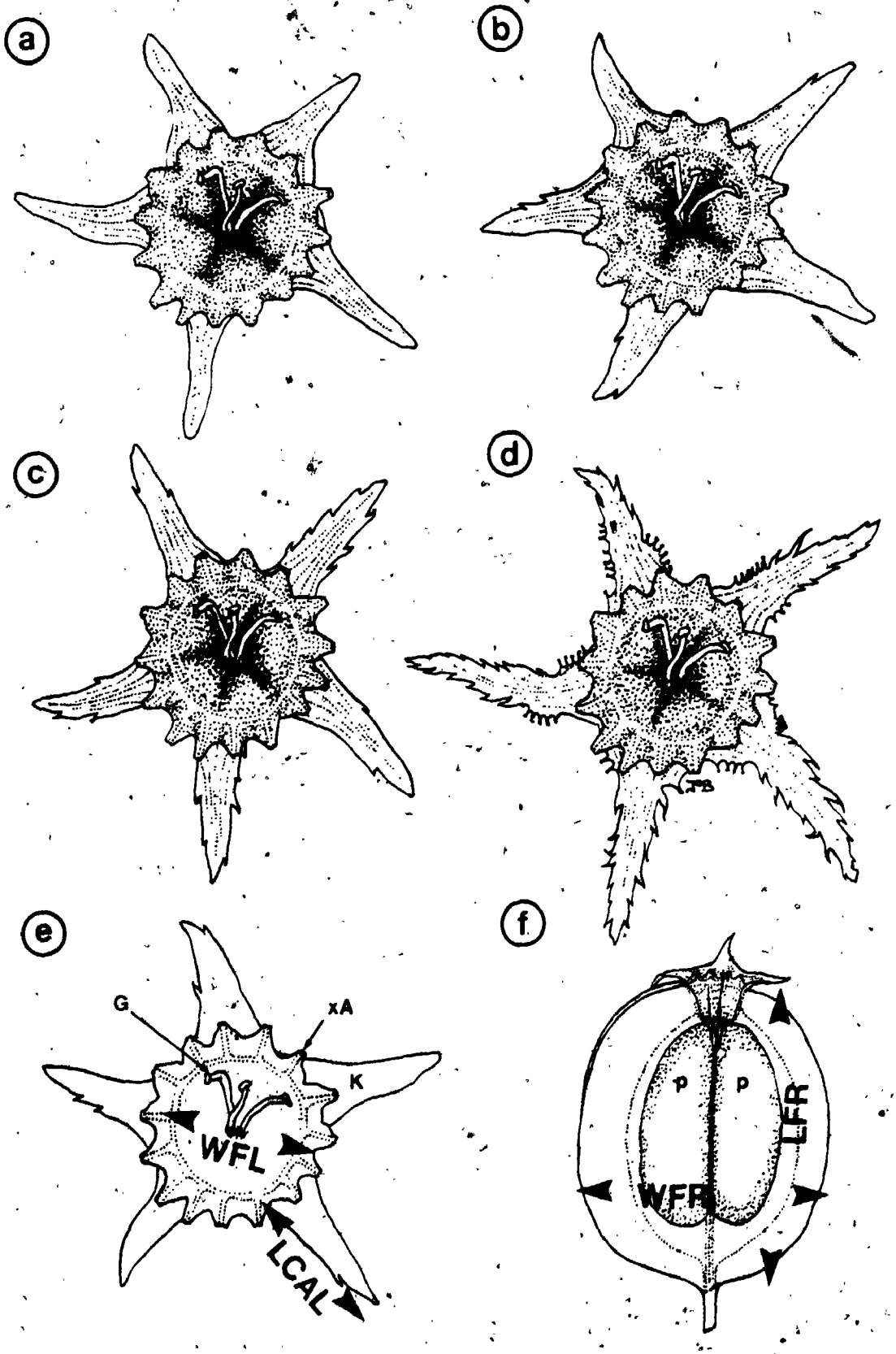

a
b
c
d
e
f
G
xA
K
WFL
LCAL
p
p
LFR
WFR

Table 3.1 Flower and fruit descriptors, showing states scored for each one. Maxima and minima observed in the 160 OTU sample are underlined.

STYL - Number of styles per flower (0, 1, 2, 3, 4, 5).

STAM - Number of stamens per flower (2-4, 5-7, 8-10, 11-13, 14-16, 17-19, 20-22, 23-25).

PROJ - Presence or absence of an apical projection of the stamen filament between the anther lobes (0, 1).

ANTH - Color of the undehisced anthers at time of anthesis (ivory, 0; faint pink, 1; pink, 2; red, 3).

TCAL - Degree of calyx lobe toothing; see Figure 3.1 for illustration of states (absent, or only one or two teeth on a single lobe, 0; isolated teeth on more than one lobe, 1; several on most lobes, 2; each lobe densely toothed, 3).

Table 3.1 Cont.

WFL - Flower width in mm (8 states at 0.5 mm intervals 2.5 - 6.5 mm; Figure 3.1).

LCAL - Length of longest calyx lobe in mm (14 states at 0.5 mm intervals 3.0 - 10.0 mm; Figure 3.1).

PUB1 - Ovary pubescence (absent, 0; sparse, 1; dense, at least locally, 2; very dense, 3).

PUB2 - Pedicel pubescence, scored as for PUB1.

LFR - Fruit length in mm (10 states at 1.15 mm intervals 7.0 - 18.5 mm; Figure 3.1).

WFR - Fruit width in mm (10 states at 1.35 mm intervals 7.5 - 21.0 mm; Figure 3.1).

3.5 Leaf Descriptors

A stratified random subsample of 60 OTUs (in which the eight strata represented the seven randomly sampled topodemes, plus the two randomly chosen C. ?grandis; Table 3.2) was drawn from the total of 113 randomly chosen OTUs in the study.

The leaf collections made for these OTUs were scored in each of two ways. Each leaf from a single short shoot was scored for descriptors X, Y and Z (Table 3.3, Fig. 3.2). The complete suite of six leaf descriptors (Table 3.3) was scored only on the terminal leaf of each of the 6-10 short shoots collected for each OTU. The terminology used here to describe leaf architecture is that of Hickey (1973).

Table 3.2 Distribution of OTUs among the strata of the 113 OTU random sample, and of its 60 OTU subsample. Strata are topodeme samples, except for the one containing the two OTUs of C. ?grandis. Symbols are those used to identify the topodeme samples in Figure 2.1 and in the illustrations of the following chapters. Sites are as identified in Figure 2.1 and Section 2.5.

(a) 113 OTU Sample.

Taxa:	Sites: 1	2	3	4	5
	T1 □	T4 ○		☆	
20-stamen C. crus-galli	10	24	0	(2*)	0
	T2 ■	T5 ●	T3 ▲	T6 ★	
10-stamen C. crus-galli	17	7	16	19	0
					T7 p
C. punctata	0	0	0	0	18

*OTUs of C. ?grandis.

Table 3.2 Cont.

(b) 60 OTU Subsample.

Taxa:	Sites: 1	2	3	4	5
20-stamen C. crus-galli	T1 8	T4 10	0	(2*)	0
10-stamen C. crus-galli	T2 8	T5 7	T3 8	T6 9	0
C. punctata	0	0	0	0	T7 8

*OTUs of C. ?grandis.

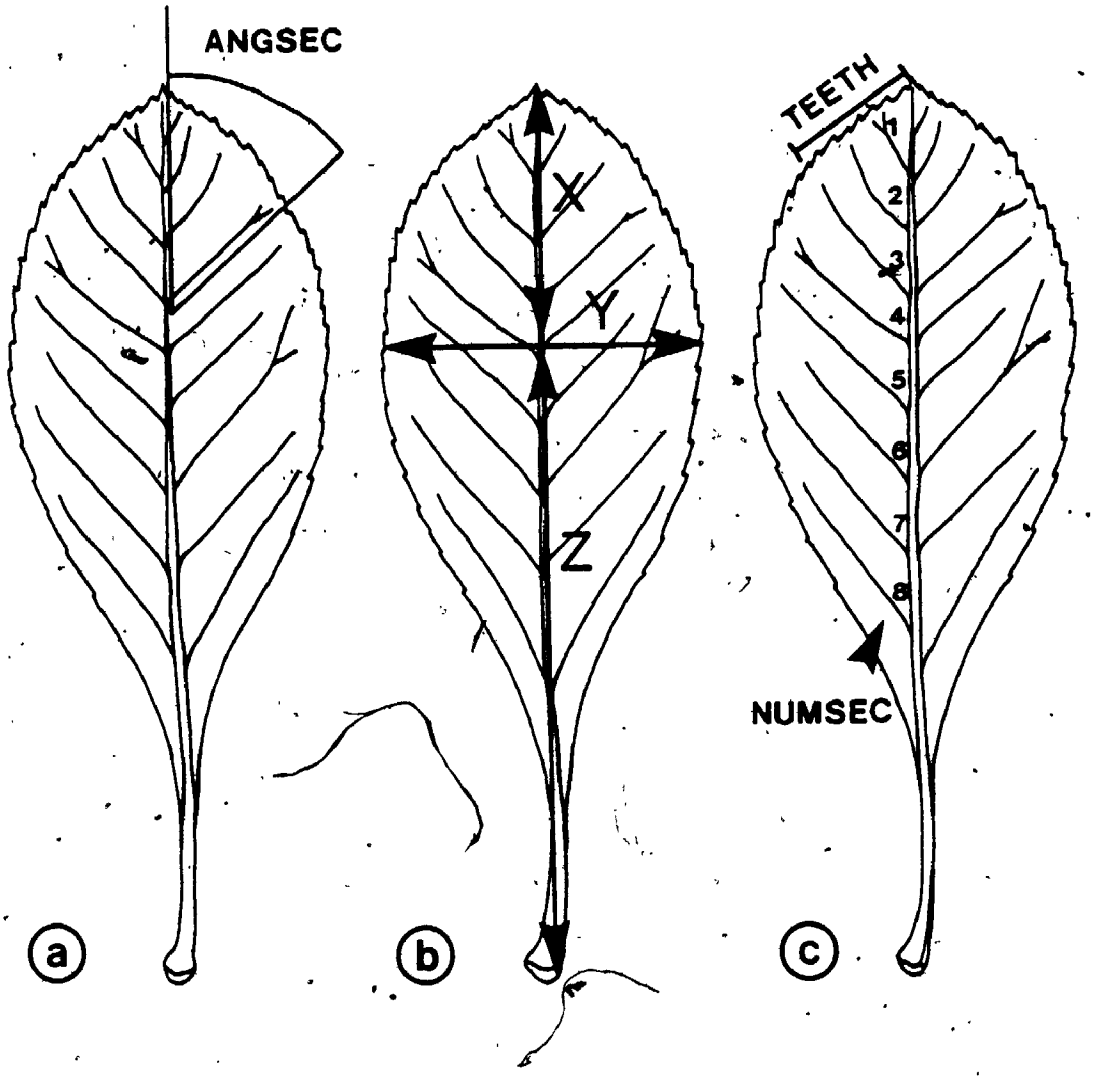


Figure 3.2 Illustrations of leaf descriptors (Table 3.3): (a) ANGSEC, angle between the midrib and the fourth secondary vein; (b) leaf dimensions X, Y and Z; (c) NUMSEC, number of secondary veins and TEETH, number of teeth per 1.0 cm on one side of the leaf apex.

Table 3.3 Leaf descriptors scored on short shoot distal leaves for each OTU in the 60 OTU subsample (Table 3.2). In parentheses, the maximum and minimum OTU means used in ranging. Refer to Figure 3.2 for further explanation.

X - Leaf length above the widest point, in mm (12.14 - 26.40).

Y - Maximum leaf width perpendicular to X and Z, in mm (14.14 - 31.20).

Z - Leaf length below widest point, to leaf base, in mm. Z and X are colinear, and are aligned between the leaf apex and the point of attachment of the leaf to the stem (29.14 - 56.14).

NUMSEC - Number of secondary veins on one side of the midrib (5.00 - 9.38).

ANGSEC - Angle between the midrib and the fourth secondary vein from the apex, to the nearest five degrees (25.00 - 43.57).

TEETH - Number of teeth per 1.0 cm of leaf margin on one side of the leaf apex (5.25 - 11.38).

3.6 Data Analysis; Summarizaton and Transformation

Data summarization and transformation was carried out by means of programs written in the BASIC and MAXBASIC languages. Use was also made of SPSS (Statistical Package for the Social Sciences, 2nd ed., Nie, Hull, Jenkins, Steinbrenner & Brent, 1975), and MINITAB (Ryan, Joiner, & Ryan, 1976, 1981), particularly for regression analyses, and for obtaining descriptive statistics for the topodeme samples (Section 3.7). Output from MINITAB was also used in preparing several of the illustrations. Further data analysis was done using programs either written for the purpose or obtained from a variety of sources cited at the relevant points in the text. Where specific sources are not cited, SPSS, MINITAB, or a hand-held electronic calculator (Texas Instruments TI-57) was used.

The raw flower and fruit data, in the form of a three-dimensional array $\underset{\sim}{X}$, with elements x_{ijk} ($i = 1...n$ OTUs, $j = 1...p$ descriptors, $k = 1...q_{ij}$ replicates) was scanned to find

$$MAX_j = \max(x_{ijk}),$$

$$\text{and } MIN_j = \min(x_{ijk}) \text{ for } j = 1...p.$$

This was done twice, once for the 113 OTU complete random samples and once for the total sample of 160 OTUs. In the same operation OTU-descriptor means were also calculated:

$$y_{ij} = (\sum_k x_{ijk})/q_{ij}.$$

Data for the short shoot terminal leaves was treated similarly, except that $MAX_j = \max(y_{ij})$ and $MIN_j = \min(y_{ij})$. Binary and multistate descriptors were treated as if they were continuous.

Data summarized in this way as a vector $\underset{\sim}{Y}_i$ of raw descriptor means for each OTU was the basis for one set of analyses in which descriptor commensurability was achieved either by standardization to zero mean and unit standard deviation ($\underset{\sim}{Y} \rightarrow \underset{\sim}{Z}$),

$$z_{ij} = (y_{ij} - \bar{y}_j)/(VAR_j)^{\frac{1}{2}},$$

where VAR_j is the variance of the n OTU means for the jth descriptor, or by ranging ($\underset{\sim}{Y} \rightarrow \underset{\sim}{R}$),

$$r_{ij} = (y_{ij} - MIN_j)/(MAX_j - MIN_j).$$

Flower and fruit data were summarized also as an array $\underset{\sim}{F}$ with elements f_{ijl} (as before, and for $l = 1 \ldots s_j$ states), where f_{ijl} is the frequency of the lth state of descriptor j for OTU i. Meristic descriptors (STYL, STAM) and continuous ones (WFL, LCAL, LFR, WFR) were converted to multistate by dissection of their global range into equal intervals. The alternative of using equally spaced frequency classes (Jardine & Sibson, 1970) or classes of equal frequency

(Blackburn, 1980) was not adopted here because of the limited sample size available for any one stratum in the study (e.g. species, morphotype, topodeme). In view of the heterogeneity of the total sample it seemed preferable to accept instead some loss of resolution as occasioned by the use of equal class intervals (Blackburn, 1980).

The number of classes used for the meristic descriptors was such that equal subdivision of the range resulted in integer values for the upper and lower bounds of each class (Table 3.1). Descriptors WFL and LCAL were scored to the nearest 0.5 mm over such a small range that they were treated as if multistate, their range being divided into 0.5 mm intervals. Fruit dimensions (LFR, WFR) were converted by subdivision of the range into ten equal parts. This number is approximately half again larger than the recommended number of classes based on the sample size available (Sturges, 1926; Blackburn, 1980). However, in view of the precision of the measurements (0.5 mm), the class widths used (1.15 and 1.35 mm) do not appear to have been excessively narrow.

3.7 Description of the Data Obtained

For the most part the methods of data analysis employed in this thesis (Chapters 4 - 8) make a number of assumptions concerning the properties of the data to be analyzed.

Univariate normality is assumed, such that replicated observations can be adequately summarized by their mean. Multivariate normality is also assumed, especially in the case of the ordination methods which rely upon eigenanalysis of dispersion matrices. Similarly, statistical inference on the basis of multivariate data requires an underlying multivariate normal distribution. The procedures used also assume that the data are continuous and linearly related to each other.

Methods for testing the distributional properties of multivariate data are available (Reyment, 1971) but may be computationally demanding (Orlóci, 1978). Gnanadesikan (1977) presents a number of graphical methods for examining multivariate data so as to detect outliers and other discontinuities, as well as departures from multivariate normality. Regrettably, it has not been possible to implement these methods for the present study. Instead, the univariate (Section 3.7.1) and bivariate (Section 3.7.2) distributions of the data have been examined. Some preliminary multivariate examinations of the data are described in Section 3.7.3. These include characterization of the individual descriptors with respect to the redundancy of the information about the sample that they provide. Section 3.7.4 reports on characteristics of the data available for each topodeme sample, including both

univariate statistics and a limited examination of their multivariate properties by means of ordinations.

3.7.1 Univariate Distributions

The distribution of each of the 17 descriptors has been examined for the random sample of 111 OTUs (60 OTUs for the leaf descriptors) by means of SPSS program CONDESCRIPTIVE and the frequency histograms and normal probability plots produced by the MINITAB package (Ryan, Joiner & Ryan, 1976, 1981). CONDESCRIPTIVE provides information on the kurtosis and skewness of descriptor distributions (Table 3.11; Nie et al, 1975). Normal probability plots are plots of the ith data point (raw OTU descriptor, mean) against the $(i - .375)/(n + .25)$ percentage point of the standard normal distribution (Ryan, Joiner & Ryan, 1976). Ties are given the same normal score, calculated from the average of their ranks. Departures from normality are indicated by non-linearity (Sokal & Rohlf, 1969, 1981; Ryan, Joiner & Ryan, 1976; Ryan & Joiner, undated), and may be described quantitatively by the correlation between the data points and their normal scores (Ryan, Joiner & Ryan, 1981; Ryan & Joiner, undated; Table 3.4).

Table 3.4 does not report on the normal plots for the binary and multistate descriptors PROJ, PUB1, PUB2 and ANTH as their distributions were, as might be expected, markedly

non-normal due to the heterogeneity of the sample. Likewise the distribution of stamen number was clearly bimodal. The departures from normality described in Table 3.4 are similarly associated principally with the meristic (STYL, TEETH) and multistate (TCAL, ANGSEC) descriptors. Only one of the remaining continuous descriptors, LCAL, also showed a significant departure from normality, due as with TCAL as well to the presence of the topodeme sample of C. punctata (T7). In the case of style number, non-normality appeared to result from the presence in the sample of OTUs with extremely low mean scores belonging to topodeme sample T6. With both ANGSEC and TEETH non-normality was due to the mixture in the sample of the two common morphotypes of C. crus-galli, since the 20-stamen crus-galli OTUs and the punctata ones behaved similarly.

3.7.2 Bivariate Distributions

Bivariate scatter plots for selected pairs of raw descriptors, and the corresponding regressions and their residuals were examined in order to check on the continuity of the data and to look for further evidence of structure in the random sample. This examination was restricted to the leaf descriptors and the six flower and fruit descriptors examined in detail in Chapters 7 and 8, namely STYL, TCAL, WFL, CAL, LFR and WFR. The basis for this selection is

evident from the preceding section, but is described in detail in Section 7.3.2.

The scatterplots indicated varying degrees of association between the descriptors, summarized below in correlation matrices for the two data sets (Table 3.6, 3.8, 3.10). More importantly, these graphs indicated the generally continuous nature of the data, and the preponderantly linear relationships between the descriptors. Athough some of the scatterplots themselves gave strong indications of underlying sample structure (particularly the leaf descriptor pairs X-TEETH, NUMSEC-TEETH and ANGSEC-TEETH), examination of standardized regression residuals and their normal scores was often even more revealing. In general the correlations between the residuals and their normal scores were quite high (> 0.97) although this is certainly below the critical value for the larger sample (N = 111). For several pairs of flower-fruit descriptors, however, the linear trend in the graph was sharply broken, the breaks corresponding to gaps between C. crus-galli s. l. and C. punctata, as well as to gaps between the 10- and 20-stamen morphotypes of C. crus-galli. For some pairs of flower-fruit descriptors (STYL-LCAL, STYL-LFR and WFL-LCAL) it was evident that the distribution of residuals was highly non-random, the non-randomness indicating contrasts between groups of topodeme samples

corresponding to the species and morphotypes in the samples. Topodeme samples T6 and T7 were especially contrasted in any comparison involving style number (STYL). Such differentiation was much less conspicuous in the regression residuals for pairs of leaf descriptors.

3.7.3 Correlation Structure

The correlation structure of the leaf, and flower and fruit data available for the random sample (ranged OTU descriptor means, R) is shown in Table 3.6 and Table 3.10. Flower and fruit descriptor correlations in the entire 160 OTU sample are shown in Table 3.8. The covariation of the ranged descriptors was also examined by means of two descriptors ranking methods described by Orlóci (1978; Table 3.5, 3.7, 3.9).

The first method, described as a "vector projection method" (Orlóci, 1973; program SPVAR, Orlóci, 1978), is based on the principle that data structure may be revealed by descriptor correlation. Accordingly, highly informative descriptors are those which are strongly correlated with other descriptors scored for the sample.

The second or "regression" method (Orlóci, 1975, 1978) is based on the opposite point of view, namely that descriptor correlation indicates redundancy, and that it is

descriptors conveying unique information which should be highly ranked. This method is based on the multiple regression of the hth descriptor on the remaining p - 1 others (h = 1...p), to obtain the corresponding squared multiple correlation coefficient, R_h^2. The variance of each descriptor can be seen as the sum of two components: one which is common to other descriptors as well, and one which is specific to the descriptor itself. R_h^2 is the ratio of the common variance to the total variance, for each descriptor (Orlóci, 1978). Accordingly R_h^2 measures the redundancy of descriptor h in terms of the extent to which its variation is predicable from that of the other descriptors.

These two methods have been used here to describe the data available (Table 3.5, 3.7, 3.9) rather than to provide a basis for the reduction of data or sampling efforts. As will be seen subsequently in Chapters 5-8, the results obtained aid in interpreting the outcomes of both classification (Chapter 5,6) and ordination (Chapter 6,8) analyses. This follows from the complementary information furnished by these two methods (Orlóci, 1978).

Rohlf (1977) has pointed out that if $\underset{\sim}{S}$ is the sample covariance matrix, not only R_h^2 may be obtained from $\underset{\sim}{S}$ and its inverse $\underset{\sim}{S}^{-1}$, but also the partial correlations among the descriptors, r_{hi}. These estimate the correlation between

descriptors h and i that would exist if the effects of the remaining p - 2 descriptors were held constant (Rohlf, 1977). The partial correlations can be compared with the product-moment correlation coefficients, r_{hi}, as a means of further elucidating descriptor relationships in the sample. In Table 3.6, 3.8 and 3.10 $|r_{hi.}|$ is usually slightly less than $|r_{hi}|$. In some cases, $|r_{hi.}| << |r_{hi}|$, indicating that the correlation between the two descriptors may be largely explained by their correlations with the other descriptors in the study. This is noticeably the case for STYL-STAM, WFL-PROJ and WFR-STAM correlations in the 160 OTU sample (Table 3.8), STYL-LCAL in the 111 OTU sample and WFL-LFR and WFL-WFR in both the 93 and 111 OTU samples (Table 3.10). Among the leaf descriptors similar behavior is seen with X-Z and Y-Z correlations. However, for four of the five correlations with TEETH, (i.e. excluding TEETH-Z) $|r_{hi.}| >> |r_{hi}|$. According to Rohlf (1977) this suggests the existence of a pattern of covariation unique to these pairs of descriptors that is obscured by covariation with the other descriptors. Such a pattern is possibly that referred to in the preceding section, in which scatterplots of TEETH with the other leaf descriptors demonstrated marked group structure in the sample.

3.7.4 Data Structure of the Topodeme Samples

Univariate statistics have been compiled for each descriptor for the individual topodeme samples as well as the samples of Crataegus ?grandis and C. sp. aff. C. bushii, and are reported in Table 3.11. These indicate that for the most part descriptor distributions within topodeme samples do not depart markedly from normality. The skewness and kurtosis of the distributions of the twelve descriptors examined in Table 3.4 were examined within each topodeme sample for which they were scored (Table 3.11). The six flower and fruit descriptors were also analyzed by means of normal probability plots (not shown). Only where (lepto-) kurtosis and skewness were extreme (H_0 of zero kurtosis, zero skewness rejected with $p < 0.001$; descriptor LFR for T2, also WFR for T2 although $0.05 > p > 0.001$; Table 3.11), due to the presence of two outlier individuals (OTUs 110, 118), did the correlation of the data with the corresponding normal scores fall below critical values ($p < 0.01$, LFR; $0.05 > p > 0.01$, WFR). The other instances of kurtosis and skewness among the topodeme samples in Table 3.11 ($0.05 > p > 0.001$) did not occasion significant departures from normality according to the correlation test. Table 3.11 also documents the way in which different descriptors (notably the binary and multistate ones PROJ, PUB1, and ANTH) are invariant within particular topodeme

samples. This consideration is of some importance to several of the analyses described in Chapters 6-8.

R-algorithm principal component analyses (see Section 6.2 for description of method) of the individual topodeme samples provided a multivariate demonstration of the homogeneity of these samples and of the absence of extreme outliers. These analyses were done using the six flower and fruit descriptors STYL, TCAL, WFL, LCAL, LFR and WFR examined in detail here and elsewhere (Chapter 7; 8). Correlations between descriptors and components indicated which of these descriptors contribute most to the scatter of each sample in its ordination space. With each sample an efficient low-dimensional summary of the data was possible, with the first component accounting for 40-55%, and the first three components together accounting for a total of 70-80% of the total variation.

3.7.5 Conclusions

The preliminary examination described here of the flower, fruit and leaf data available for the random sample suggests that the individual leaf descriptors and at least a subset of the flower and fruit ones have tolerably close to normal distributions (Table 3.4). Examination of bivariate distributions suggests further that the joint distributions of these better behaved descriptors are reasonably

continuous and not markedly non-linear. To the extent that departures from normality and continuity are observable, these appear to be due to the heterogeneity of the sample. Finally, the univariate statistics for the individual topodeme samples indicate that, within these samples too, departures from normality are infrequent and not too severe. These results by no means constitute a demonstration of multivariate normality or continuity, but they do give some reassurance at least on the former score. This is based on inferences from the central limit theorem to the effect that as individual descriptors scored for a random sample are normally distributed, so their joint distribution tends toward multivariate normality (Orlóci, 1978; Pimentel, 1979).

The multivariate analyses described in the following five chapters prove to be much more efficient than the methods used here in revealing and documenting the structure of the sample at hand. However, the results arrived at here (Section 3.7.3) already provide an indication of how the results of those chapters were obtained.

Table 3.4 Correlations between raw OTU, descriptor means and their normal probability scores, (a-b) for the random samples indicated. H_0: r = 1.0 (normality) rejected for the correlations below the critical values calculated according to Ryan and Joiner (undated) and tabulated in (c) below.

(a) Flower and fruit descriptors, 111 OTUs.

STYL	0.975	**
TCAL	0.971	**
WFL	0.985	*
LCAL	0.973	**
LFR	0.991	ns
WFR	0.984	*

(b) Leaf descriptors, 58 OTUs.

X	0.981	ns
Y	0.991	ns
Z	0.980	ns
NUMSEC	0.993	ns
ANGSEC	0.974	**
TEETH	0.955	**

ns Departure from normality not significant, $p > 0.05$.

* Departure from normality significant, $0.05 > p > 0.01$.

** Departure from normality significant, $p < 0.01$.

Table 3.4 Cont,

(c) Approximate critical values for the correlations between variates and their normal scores for the sample sizes (N) indicated.

	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
N = 58	0.983	0.979	0.970
N = 111	0.991	0.989	0.982

Table 3.5 Ranking of 6 ranged leaf descriptors (a) according to their correlation with other descriptors (Orlóci, 1973, 1978) and (b) inversely with respect to their redundancy (Orlóci, 1975, 1978; Rohlf, 1977), for the 60 OTU subsample. See text for details. See Table 3.3 for explanation of descriptor abbreviations.

	(a)		(b)	
Des (1)	% Sum of Squares (2)	Rank	% Redundancy (3)	Rank
X	8.58	4	61.79	3
Y	2.57	6	63.56	4
Z	33.43	1	67.87	5
NUMSEC	29.27	2	58.18	2
ANGSEC	6.25	5	72.01	6
TEETH	23.31	3	51.76	1

(1) Descriptor.

(2) Proportion of the trace of the (residual) dispersion matrix accounted for by the descriptor specific sum of squares.

(3) % Redundancy = $100R_h^2 = 100(1 - 1/(S_{hh}S_{hh}^{-1}))$.

Table 3.6 Semi-matrices of correlation coefficients, r_{hi} (above the principal diagonal), and partial correlations, $r_{hi.}$ (below the principal diagonal), between 6 leaf descriptors scored for the 60 OTU subsample. Partial correlations are calculated according to Rohlf (1977) and Orlóci (1978). See text for details. See Table 3.3 for explanation of descriptor abbreviations.

X	-	0.562	0.491	-0.065	-0.050	-0.348
Y	0.500	-	0.616	0.155	-0.073	0.244
Z	0.279	0.426	-	-0.210	-0.552	0.212
NUMSEC	-0.048	0.057	0.164	-	0.672	0.134
ANGSEC	0.045	0.232	-0.559	0.695	-	-0.257
TEETH	-0.581	0.436	0.036	0.302	-0.341	-
	X	Y	Z	NUMSEC	ANGSEC	TEETH

Table 3.7 Ranking of 11 ranged flower and fruit descriptors (a) according to their correlation with other descriptors (Orlóci, 1973, 1978) and (b) inversely with respect to their redundancy (Orlóci, 1975, 1978; Rohlf, 1977), for the 160 OTU sample. See text for details. See Table 3.1 for explanation of descriptor abbreviations.

	(a)		(b)	
Des (1)	% Sum of Squares (2)	Rank	% Redundancy (3)	Rank
STYL	22.93	1	86.75	9
STAM	10.35	2	88.02	11
PROJ	- (4)	8	75.68	6
TCAL	8.72	7	49.48	3
PUB1	28.55	3	87.50	10
WFL	- (4)	9	74.37	4
LCAL	19.04	4	43.87	1
LFR	- (4)	10	75.03	5
WFR	118.46	6	83.60	8
PUB2	- (4)	11	79.90	7
ANTH	36.50	5	46.23	2

Table 3.7 Cont.

(1) Descriptor.

(2) Proportion of the trace of the (residual) dispersion matrix accounted for by the descriptor specific sum of squares.

(3) % Redundancy = $100R_h^2 = 100(1 - 1/(S_{hh}S_{hh}^{-1}))$.

(4) Trace of the residual dispersion matrix < 0.

Table 3.8 Semi-matrices of correlation coefficients, r_{hi} (above the principal diagonal), and partial correlations, $r_{hi.}$ (below the principal diagonal), between 11 flower and fruit descriptors scored for the entire 160 OTU sample. Partial correlations calculated according to Rohlf (1977) and Orlóci (1978). See text for details. See Table 3.1 for explanation of descriptor abbreviations.

STYL	-	0.748	0.393	0.278	0.731	0.585	0.386	0.274	0.608	0.680	0.122
STAM	0.199	-	0.731	-0.002	0.500	0.740	0.210	0.469	0.712	0.330	0.253
PROJ	0.165	0.660	-	-0.146	0.012	0.575	0.034	0.227	0.453	-0.117	0,011
TCAL	0.311	0.015	-0.302	-	0.150	0.194	0.380	-0.332	-0.074	0.247	0.031
PUB1	0.434	0.383	-0.343	-0.253	-	0.230	0.484	0.251	0.373	0.853	0.416
WFL	0.215	0.332	-0.052	0.228	-0.243	-	0.293	0.434	0.708	0.139	0.238
LCAL	-0.125	-0.266	0.222	0.268	0.304	0.209	-	0.048	0.238	0.460	0.316
LFR	-0.134	0.091	-0.216	-0.300	0.068	-0.027	-0.097	-	0.773	0.152	0.364
WFR	0.245	0.102	0.017	-0.063	-0.217	0.285	0.143	0.704	-	0.320	0.265
PUB2	0.265	-0.111	-0.148	0.065	0.507	-0.165	0.069	-0.094	0.190	-	0.266
ANTH	-0.443	0.065	0.048	0.172	0.401	0.201	0.090	0.216	-0.047	-0.012	-
	STYL	STAM	PROJ	TCAL	PUB1	WFL	LCAL	LFR	WFR	PUB2	ANTH

Table 3.9 Ranking of 6 ranged flower and fruit descriptors (a) according to their correlation with other descriptors (Orlóci, 1973, 1978) and (b) inversely with respect to their redundancy (Orlóci, 1973, 1978; Rohlf, 1977), for the 93 OTU (T1 - T6) and 111 OTU (T1 - T7) random samples. See text for details. See Table 3.3 for explanation of descriptor abbreviations.

(a) Correlation.

	93 OTUs		111 OTUs	
Des (1)	% SS (2)	Rank	% SS (2)	Rank
STYL	65.31	3	11.14	3
TCAL	32.51	5	25.32	2
WFL	70.07	4	7.72	5
LCAL	- (3)	6	9.10	4
LFR	21.71	2	2.25	6
WFR	46.32	1	45.71	1

Table 3.9 Cont.

(b) Redundancy.

	93 OTUs		111 OTUs	
Des (1)	% Red (4)	Rank	% Red (4)	Rank
STYL	69.63	4	72.76	5
TCAL	37.18	2	49.22	2
WFL	62.25	3	65.04	4
LCAL	17.18	1	20.37	1
LFR	70.86	5	56.84	3
WFR	80.37	6	75.93	6

(1) Descriptor.

(2) Proportion of the trace of the (residual) dispersion matrix accounted for by the descriptor specific sum of squares.

(3) Trace of residual dispersion matrix < 0.

(4) % Redundancy = $100R_h^2 = 100(1 - 1/(S_{hh}S^{-1}_{hh}))$.

Table 3.10 Semi-matrices of correlation coefficients, r_{hi} (above the principal diagonal), and partial correlations, $r_{hi.}$ (below the principal diagonal), between 6 flower and fruit descriptors scored for (a) 93 OTUs (T1 - T6) and (b) 111 OTUs (T1 - T7). Partial correlations calculated according to Rohlf (1977) and Orlóci (1978). See text for details, and Table 3.1 for explanation of descriptor abbreviations.

(a) 93 OTUs (10- and 20-stamen C. crus-galli only)

STYL	-	0.252	0.665	0.049	0.239	0.686
TCAL	0.326	-	-0.108	0.160	-0.460	-0.113
WFL	0.425	-0.171	-	0.105	0.515	0.732
LCAL	-0.243	0.074	0.175	-	-0.185	0.073
LFR	-0.362	-0.336	0.162	-0.348	-	0.698
WFR	0.564	0.037	0.183	0.281	0.678	-
	STYL	TCAL	WFL	LCAL	LFR	WFR

(b) 111 OTUs (C. crus-galli plus C. punctata)

STYL	-	0.559	0.676	0.337	0.311	0.669
TCAL	0.575	-	0.145	0.407	-0.129	0.147
WFL	0.406	-0.242	-	0.226	0.532	0.753
LCAL	0.036	0.269	0.147	-	-0.032	0.142
LFR	-0.153	-0.143	0.091	-0.065	-	0.707
WFR	0.395	-0.054	0.334	-0.029	0.587	-
	STYL	TCAL	WFL	LCAL	LFR	WFR

Table 3.11 Descriptive statistics for 11 flower and fruit descriptors and 6 descriptors of short shoot distal leaves, for six C. crus-galli (T1-6) and one C. punctata (T7) topodeme samples, plus samples of C. ?grandis (Site 4) and C. sp. aff. C. bushii (Site 6). Statistics shown are (1) the mean and (2) the range. The ranges shown represent the absolute maxima and minima observed for each group, and are tabulated only for the continuous and meristic flower and fruit descriptors. The ranges of the leaf descriptors as well as all the means are calculated for each group from OTU means, however. In addition, the results of testing the significance of descriptor skewness and kurtosis within each group are also given (a - f), for the descriptors whose ranges are tabulated. See Table 3.1 and 3.3 for explanation of descriptors.

Table 3.11 Cont.

Sample	T1	T2	T3	T4	T5	T6	T7	?grandis	cf. bushii
N	10	17	16	24	7	19	18	7	11
STYL (Style number)									
(1)	2.10 a,e	1.58	1.32	2.25	1.95	1.04	3.11	2.63	2.00 c
(2)	1.0-3.0	1.0-3.0	1.0-3.0	1.0-4.0	1.0-3.0	0-2.0	2.0-5.0	2.0-4.0	1.0-3.0
STAM (Stamen number)									
(1)	17.18 e	8.56	7.12	18.45	10.65	9.37	19.08 a	19.42	11.56
(2)	12.0-20.0	4.0-13.0	4.0-11.0	13.0-21.0	7.0-17.0	5.0-13.0	13.0-23.0	15.0-21.0	9.0-19.0
PROJ (Stamen projection)									
(1)	0.94	0.03	0 *	0.98	0 *	0 *	0.40	0.52	0.01
ANTH (Anther color)									
(1)	0.20	0 *	0 *	0.91	0 *	3.00 *	2.28	3.00 *	1.64
TCAL (Calyx lobe toothing)									
(1)	0.81	0.96 a,e	0.84	0.75	0.44	0.40	1.29	0.40	3.00 b,f
(2)	0-2.0	0-2.0	0-2.0	0-2.0	0-1.0	0-2.0	0-3.0	0-2.0	2.0-3.0
PUBl (Ovary pubescence)									
(1)	0 *	0.22	0 *	0.01	0.02	0 *	2.97	2.79	0.01

Table 3.11 Cont.

Sample	T1	T2	T3	T4	T5	T6	T7	?grandis	cf. bushii
N	10	17	16	24	7	19	18	7	11
PUB2 (Pedicel pubescence)									
(1)	0.20	1.25	0 *	0.09	0.71	0.11	3.00 *	2.57	0.90
WFL (Flower width, mm)									
(1)	4.24	3.40	3.54	4.65 f	4.00	3.87	4.47	4.14	4.48
(2)	3.5-5.0	2.5-4.0	2.5-5.0	4.0-5.5	3.0-4.5	3.5-4.5	3.5-5.0	3.0-5.0	3.0-5.5
LCAL (Length, longest calyx lobe, mm)									
(1)	4.29	4.83	4.67	4.76	4.46	4.65	5.20	6.44	5.36
(2)	3.0-5.5	3.5-6.0	4.0-6.0	3.5-6.5	3.5-5.5	3.5-6.5	3.5-8.5	4.5-7.5	4.0-7.5
LFR (Fruit length, mm)									
(1)	11.27	10.48 d,f	10.58	11.90 e	12.88	11.86	12.09	13.91	11.23
(2)	9.0-14.5	8.0-14.0	8.5-14.0	8.5-15.0	10.0-15.5	9.0-15.0	8.0-16.0	10.0-18.5	8.5-13.0
WFR (Fruit width, mm)									
(1)	11.13	11.10 e	9.97	13.75	13.26	11.16	13.35	15.07	12.37
(2)	8.5-14.5	8.0-15.0	7.5-13.5	10.0-18.0	10.0-16.0	7.5-14.0	9.5-17.0	10.5-21.0	8.0-16.0

Table 3.11 Cont.

Sample	T1	T2	T3	T4	T5	T6	T7
N	10	17	16	24	7	19	18
X (Leaf length above widest point, mm)							
(1)	20.08	17.17	15.11	18.37	21.78 a,e	19.60	19.25
(2)	16.3-23.2	16.0-19.0	12.1-17.3	15.2-20.8	18.4-22.9	16.6-22.5	16.6-21.6
Y (Leaf width at widest point, mm)							
(1)	18.68	18.10	19.14	21.84	20.89 a	21.33	22.90
(2)	16.1-21.5	15.7-20.0	14.1-22.6	19.5-25.0	18.3-22.1	18.4-23.4	19.0-28.3
Z (Leaf length below widest point, mm)							
(1)	43.38	40.11	35.36	41.73	37.27	39.13	45.28
(2)	35.9-48.3	34.3-44.4	29.1-38.0	37.0-49.7	34.6-40.9	35.6-43.0	40.0-56.1
NUMSEC (Number of secondary veins)							
(1)	5.97	7.08	8.31	6.22	7.25	7.92	7.07
(2)	5.6-6.4	6.6-7.6	6.9-9.4	5.0-7.0	6.6-8.2	7.0-8.7	6.2-7.6
ANGSEC (Angle between midrib and fourth secondary vein from leaf apex)							
(1)	28.40	33.25	38.55	30.08	37.76	39.20	29.42 a,e
(2)	25.0-31.3	30.0-37.3	35.7-40.0	27.9-32.1	31.4-40.0	36.3-43.6	25.6-31.4

Table 3.11 Cont.

Sample	T1	T2	T3	T4	T5	T6	T7
N	10	17	16	24	7	19	18
TEETH (Number of teeth in 1.0 cm of margin to one side of leaf apex)							
(1)	6.11	7.07	7.85	7.38	5.81	6.66 a	9.94
(2)	5.3-6.8	6.4-7.7	7.1-8.6	6.0-8.4	5.4-6.1	5.9-7.0	8.8-11.2

* Descriptor indicated is invariant within group (variance = zero).

(a) Distribution skewed to left ($p < 0.05$).

(b) Distribution skewed to left ($p < 0.001$).

(c) Distribution skewed to right ($p < 0.05$).

(d) Distribution skewed to right ($p < 0.001$).

(e) Distribution leptokurtic ($p < 0.05$).

(f) Distribution leptokurtic ($p < 0.001$).

The significance level given for (a - f) above is that at which H_0 of zero skewness or zero kurtosis is rejected. For all descriptors for which ranges (2) are tabulated, where no letters appear next to the mean (1), skewness and kurtosis are not significantly different from zero.

CHAPTER FOUR

RESEMBLANCE MEASURES

4.1 Introduction

One of the major goals of this thesis is to determine the group structure of a large and geographically representative sample of Ontario Crataegus crus-galli L. sensu lato. This is done in order to have a basis for comparison with existing reference classifications of Crataegus. By group structure is meant the differentiation of polythetically defined groups of similar OTUs from one another (Cormack, 1971). Accordingly, one means of describing this differentiation is to sort OTUs into groups on the basis of their resemblance to one another. Formal methods for sorting are discussed in the next chapter; the present chapter is concerned with the methods used to assess between-OTU resemblance. This assessment may be seen as a translation from measurement space ($\underset{\sim}{X}$) to resemblance space (-s) of information descriptive of the OTUs (Fig. 1.1).

A range of potentially useful resemblance measures is available. Selection is largely a matter of matching the kind of data on hand with an appropriate measure. Resemblance measures must also be compatible with the

subsequent analyses that are intended. Since different measures may assess different aspects of the resemblance between OTUs it may be useful to employ more than one, and to compare the results, either directly (Section 4.5) or by means of the classifications which they produce (Section 5.4, Section 6.3).

In the present case, of the 17 descriptors in use (Table 3.1, 3.3), ten are continuous or meristic and seven are ordered multistate. All but two (ANTH, PUB2) of the latter are scored on several replicates for each OTU. With the exception of these two, all the descriptors are summarized as either a mean in data matrix $\underset{\sim}{Y}$ or as an observed frequency distribution (in $\underset{\sim}{F}$). Three related resemblance measures have been selected for use here. Two of these may be calculated from $\underset{\sim}{Y}$ or one of its transforms, $\underset{\sim}{R}$ (ranged means) or $\underset{\sim}{Z}$ (standardized means), while the third is calculated from $\underset{\sim}{F}$. A total of four between-OTU resemblance matrices have been calculated in this way for the entire 160 OTU sample.

4.2 Euclidean Distance

This measure is attractive because of its intuitive simplicity. Distances between objects defined by p-element vectors of descriptor means are calculated as

$$e(h,i) = \left(\sum_j (y_{hj} - y_{ij})^2\right)^{\frac{1}{2}}, \quad \text{for } h = 2...n,$$
$$i = 1...(h-1),$$
$$j = 1...p.$$

However, euclidean distances calculated in this way are meaningless if the p descriptors used are not commensurate (Orlóci, 1978). Since the raw descriptors employed are evidently not commensurate (Table 3.1, 3.3; $\underset{\sim}{X}$, $\underset{\sim}{Y}$), this requirement is satisfied by calculating distances between OTUs from either $\underset{\sim}{R}$ or $\underset{\sim}{Z}$, as follows:

$$e(h,i) = \left(\sum_j (r_{hj} - r_{ij})^2\right)^{\frac{1}{2}}$$

or $$*e(h,i) = \left(\sum_j (z_{hj} - z_{ij})^2\right)^{\frac{1}{2}}, \quad \text{for } h = 2...n,$$
$$i = 1...(h-1),$$
$$j = 1...p.$$

In both cases the calculations were performed using a modification of program EUCD (Orlóci, 1978), or Podani's NCLA program (Podani, 1980).

Finally, calculation of euclidean distances between OTUs is affected by the correlations betweem the descriptors used. Only if these correlations are zero will the distances be comparable with those calculated for a different set of OTUs (in which the same descriptors are similarly uncorrelated). Moreover, even in the study of a single set of data, distances between OTUs may be distorted by descriptor correlations (Blackith & Reyment, 1971). If,

as here, the purpose of the study is to describe the relative affinities of OTUs then it may be that such distortion is to be avoided, if possible. One method of doing so is discussed in the following two sections.

4.3 Generalized Distance

Mahalanobis' (1936) generalized distance between objects may be described as the distance that obtains in a space in which the descriptors of the OTUs are uncorrelated. This space is described by a "fundamental tensor", the inverse of the sample dispersion matrix (Blackith & Reyment, 1971). Thus the generalized distance is formulated as

$$m(h,i) = ((\underset{\sim}{x}_h - \underset{\sim}{x}_i)'\underset{\sim}{S}^{-1}(\underset{\sim}{x}_h - \underset{\sim}{x}_i)) ,$$

where $\underset{\sim}{x}_h$ and $\underset{\sim}{x}_i$ are vectors of observations describing objects h and i respectively.

4.3.1 Among OTUs

Generalized distances among OTUs based on the inverse of the 160 OTU sample dispersion matrix were obtained by a computationally simpler formulation suggested by Bradfield and Orlóci (1975; Orlóci, 1978). They point out that an equivalent to the expression for m(h,i) given above can be written in terms of the component scores (w_{hk}, w_{ik}) and eigenvalues (λ_k) produced by the R-algorithm of principal

components analysis (PCA; Section 6.2). Accordingly

$$m(h,i) = (\sum_k (w_{hk} - w_{ik})^2/\lambda_k)^{\frac{1}{2}}, \quad \text{for } k = 1...t \text{ components.}$$

Distances were calculated using an adaptation of program EUCD, after first using program PCAR to calculate eigenvalues and component scores (Orlóci, 1978). Identical results were obtained whether the PCA was based on the covariance matrix for the OTU vectors in $\underset{\sim}{R}$, or the corresponding correlation matrix for $\underset{\sim}{Y}$.

4.3.2 Among Topodeme Samples

Generalized distances between topodeme samples are used in Section 8.6 to evaluate the results of canonical variates analyses. Generalized distances were calculated between topodeme samples h and i using their mean vectors and the inverse of their joint covariance matrix, $\underset{\sim}{S}_{hi}^{-1}$. Joint covariance matrices were calculated in two ways as described below, from all pairs of topodeme sample covariance matrices ($\underset{\sim}{S}_h$, $\underset{\sim}{S}_i$) or from p x n matrices of ranged OTU descriptor means ($\underset{\sim}{R}'_h$, $\underset{\sim}{R}'_i$) for h = 2...n, i = 1...(h - 1).

4.3.2.1 Pooled Covariance Matrix

A pooled covariance matrix $\underset{\sim}{S}_{hi}$ (Orlóci, 1978) was calculated from $\underset{\sim}{S}_h$ and $\underset{\sim}{S}_i$, with elements

$$s_{jk} = [(n_h - 1)s_{hjk} + (n_i - 1)s_{ijk}]/(n_h + n_i - 2),$$

for j and k = 1...p.

4.3.2.2 Common Covariance Matrix

Where n_h and n_i are unequal ($n_h < n_i$, say) Anderson (1958) suggests the following calculation of a common covariance matrix $\underset{\sim}{S}_{hi}$ in which $(n_i - n_h)$ OTU vectors in the larger data set ($\underset{\sim}{R}_i$) are ignored. This approach assumes that the sequence of OTU vectors in each data set is independent of the contents of those vectors.

$$\underset{\sim}{S}_{hi} = \sum_a [\underset{\sim}{N}][\underset{\sim}{N}]',$$

where $\underset{\sim}{N} = [\underset{\sim}{r}_{ha} - \underset{\sim}{\bar{r}}_h - (n_h/n_i)^{\frac{1}{2}} (\underset{\sim}{r}_{ia} - (1/n_h)(\sum_b \underset{\sim}{r}_{ib}))]$,

for a = 1...n_h, and b = 1...n_h.

4.4 Information Radius

The formulation of the generalized distance given above assumes a multivariate normal distribution of the data as well as, when calculated from joint covariance matrices, the equality of the individual group covariance matrices. The information radius (Sibson, 1969; Jardine & Sibson, 1971) represents a generalization of m(h,i) to situations in which descriptor state distributions for individual OTUs may be of

any kind (Jardine 1971).

In addition, the information radius was employed in the present study as a means of taking into account an unknown amount of within-OTU variability in descriptor expression. Such variability has been dealt with by numerical taxonomists in a a number of ways. Resemblance coefficients which weight differences between OTU mean vectors inversely according to the within-OTU variances of the individual elements (descriptors) have been proposed by Pearson (1926), Sanghvi (1953), Crovello (1968) and Flake, v. Rudloff and Turner (1969). Limitations of the first three of these methods have been discussed by Blackburn (1980). McNeill (1974) proposed a method based on composite multistate descriptors in which the secondary (for multistate descriptors) and tertiary (for continuous ones) descriptor states represent the entire range of values possible. Three levels of descriptors are used in the case of continuous descriptors since two methods of representing their variability were employed. Estimates of resemblance are thus based on the degree of overlap or non-overlap of OTU descriptor state distributions. Blackburn (1980) has proposed a resemblance function based on comparisons of descriptor ranges for pairs of OTUs. A more simplistic method for doing the same thing has also been proposed by Sinnott (1981).

Alternatively, if a within-OTU covariance matrix is meaningful, the generalized distance described above may be used to scale between-OTU resemblances according to OTU variability. The generalized distance to a relatively invariant OTU will be large, but will tend to decrease as OTUs become increasingly variable, in a manner related to the probability of successful OTU discrimination.

The information radius takes into account variability in OTU-descriptor state expression by reducing the distance between OTUs according to the degree of overlap of their descriptor-state frequency distributions. It was selected for use here because of its independence from assumptions concerning the underlying distributional properties of the data. Descriptor state frequencies for OTUs were estimated from the data available, as described in Section 3.6. In the case of continuous descriptors an alternative would have been to sum the values of the univariate information radius tabulated by Jardine and Sibson (1971), under an assumption of underlying normal distributions for each descriptor for each OTU pair, as described by Prentice (1979). In view of the sample size available for each OTU, ten replicates each for WFL and LCAL, and twenty for the rest (STAM, LFR and WFR) it was felt that the method used was likely to be adequate. Prentice (1979) suggests that the effect of small sample size (overestimation of the information radius) in

any case will be about constant for any one data set.

The information radius between pairs of OTUs was calculated according to the formulation given by Prentice (1979) from the observed descriptor-state frequencies in $\underset{\sim}{F}$, as follows:

$$d(h,i) = \sum_j 0.5 \sum_l (f_{hl} \log_2 (2f_{hl}/(f_{hl} + f_{il})) + f_{il} \log_2 (2f_{il}/(f_{hl} + f_{il}))),$$

for h = 2...n and i = 1...(h-1) OTUs, l = 1...s_j descriptor states, and j = 1...p descriptors.

Information radii were also calculated between topodeme samples in a manner analoguous to that described for m(h,i) in Section 4.3.2 above, as part of a check on the performance of canonical variates analyses (Section 8.6).

Orlóci (1978) points out that the information radius represents the average divergence of the two OTU descriptor-state frequency distributions from an equidistribution whose elements are the mean frequencies of each descriptor-state for the two OTUs. He goes on to identify this measure as one of a larger class of minimum discrimination information statistics.

4.5 Comparison of Resemblance Matrices

Resemblance matrices may be compared by examining the correlations among them over the corresponding $n(n - 1)/2$ unique elements in each matrix (Rohlf & Sokal, 1981; Podani, 1982). The correlation matrix resulting from such a comparision may then serve for both ordination and classification of the resemblance matrices.

An alternative method of comparing resemblance matrices has been described by Lefkovitch (1978; Small, Lefkovitch & Classen, 1982), but has not been implemented here. The method is based on calculating coordinates for OTUs from each resemblance matrix (Gower, 1971b) and then finding the consensus configuration of OTUs (Lefkovitch, 1978). Squared distances (Gower, 1971b) between data sets represented by resemblance matrices and their consensus can then be analyzed to determine the congruence of the data sets (Small, Lefkovitch & Classen, 1982). Lefkovitch (1982) has also suggested a non-parametric test for evaluating the relationship among resemblance matrices. These methods are of interest since they avoid the principal statistical shortcoming of the method employed here, namely the lack of independence among the distances contributing to the correlation coefficients which are calculated (Table 4.1, 4.3).

4.5.1 160 OTU Sample

The correlations among the four resemblance matrices (**E**, ***E**, **M**, **D**) calculated from the flower and fruit data for the entire 160 OTU sample are all quite high (Table 4.1). However, the matrix of generalized distances, **M**, is evidently the most dissimilar, having the lowest correlations in the matrix with the other three resemblance matrices (Table 4.1).

Ordination of the four resemblance matrices by means of Q-algorithm PCA (Table 4.1, Section 6.2) reflects their high degree of similarity. All four resemblance matrices have positive scores on the first principal component of their correlation matrix. This component also accounts for 79.3% of the trace. In view of the limited comparisons possible among only four objects, further discussion of these results will be deferred to Section 5.5, where they can be interpreted in light of the classifications to which they give rise.

4.5.2 60 OTU Sample

As described in Section 3.5 and 3.6, a subsample of 60 OTUs was scored for leaf as well as flower and fruit descriptors. The basic data matrix for these analyses was a 60 x 17 array of OTU descriptor means, each descriptor

transformed to a 0 - 1.0 range as described for the leaf data in Section 3.6. This was done to avoid the unequal weighting of binary and continuous descriptors versus continuous and meristic ones described in Section 5.4.4 below for the 160 OTU sample. Ranged in this way, the euclidean distance calculated from **R** is the same as the "Euclidean distance for mixed data" proposed by Podani (1980) as an analogue of Gower's coefficient S_G (Gower, 1971a).

In subdividing the 17 descriptors into four contrasting data sets (Table 4.2) two descriptors, style number (STYL) and stamen number (STAM), were excluded entirely. This was done because it was felt that these descriptors by themselves determined one possible classification of the sample (see Section 5.4.2, below), and that their removal would enable comparison of the group structures implied by the other descriptors. The complete set of all 17 descriptors provided a fifth data set. In addition to matrices of euclidean distances, matrices of generalized distances were also calculated among OTUs for each data set. The matrix of information radii calculated for the 60 OTU subsample from the observed OTU descriptor-state frequencies for the eleven flower and fruit descriptors provided a reference matrix. In this way a total of m = 11 different resemblance matrices (Table 4.2) were available for comparison, as described above.

Classification of the resemblance matrices (Figure 4.1) on the basis of their intercorrelations shows a marked contrast between those based on the flower and fruit data and those based on the leaf data.

Ordination of the resemblance matrices (Q-PCA; Table 4.3; Fig. 4.2) illustrates more clearly the trends in the comparison, however. Scores on the first principal component, which accounts for 70.7% of the trace of the correlation matrix, are all positive and little scattered (Table 4.3). This is a reflection of the high degree of similarity among the eleven matrices (Podani, personal communication). The second and third principal components, although together they account for only 17.4% of the trace nevertheless reveal more about the relationships between the resemblance matrices. The second principal component appears to represent a trend from reproductive and multistate data to continuous and vegetative data, while the third principal component appears to represent a contrast between euclidean and generalized distance matrices (Fig. 4.2). Euclidean distances based on all 17 descriptors, or on 9 reproductive ones, or 7 multistate ones (Table 4.2) were most similar to the information radius reference matrix (based on all 11 reproductive descriptors; Table 4.2; Fig. 4.1, 4.2). More detailed discussion of the 60 OTU results is deferred to Chapter 6 (Section 6.3).

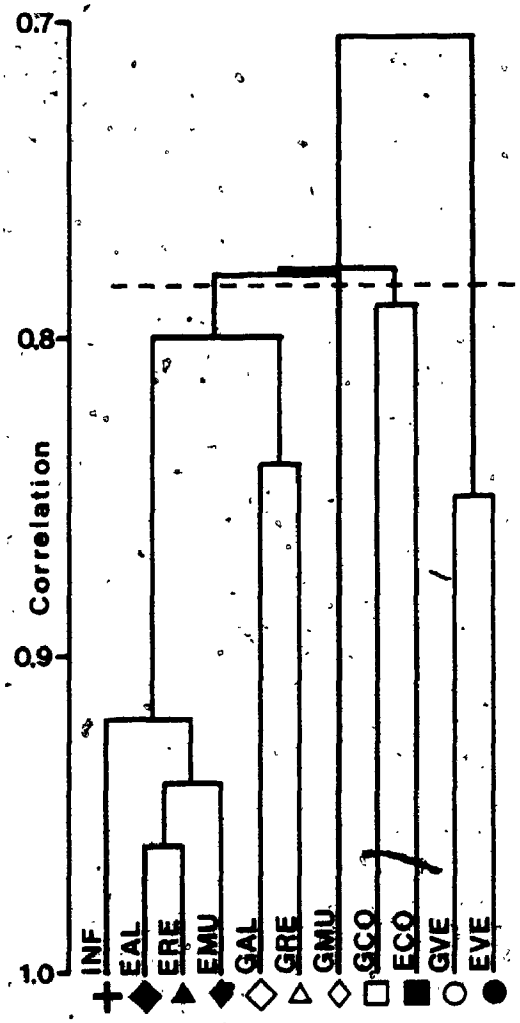


Figure 4.1 Single-linkage clustering of eleven resemblance matrices, based on their intercorrelations (Table 4.3). See Table 4.2 for explanation of the abbreviations of the matrices. Scale at left represents the maximum correlation level at which the matrices in any given group were joined. Dashed line indicates level at which clusters shown in Figure 4.2 are formed.

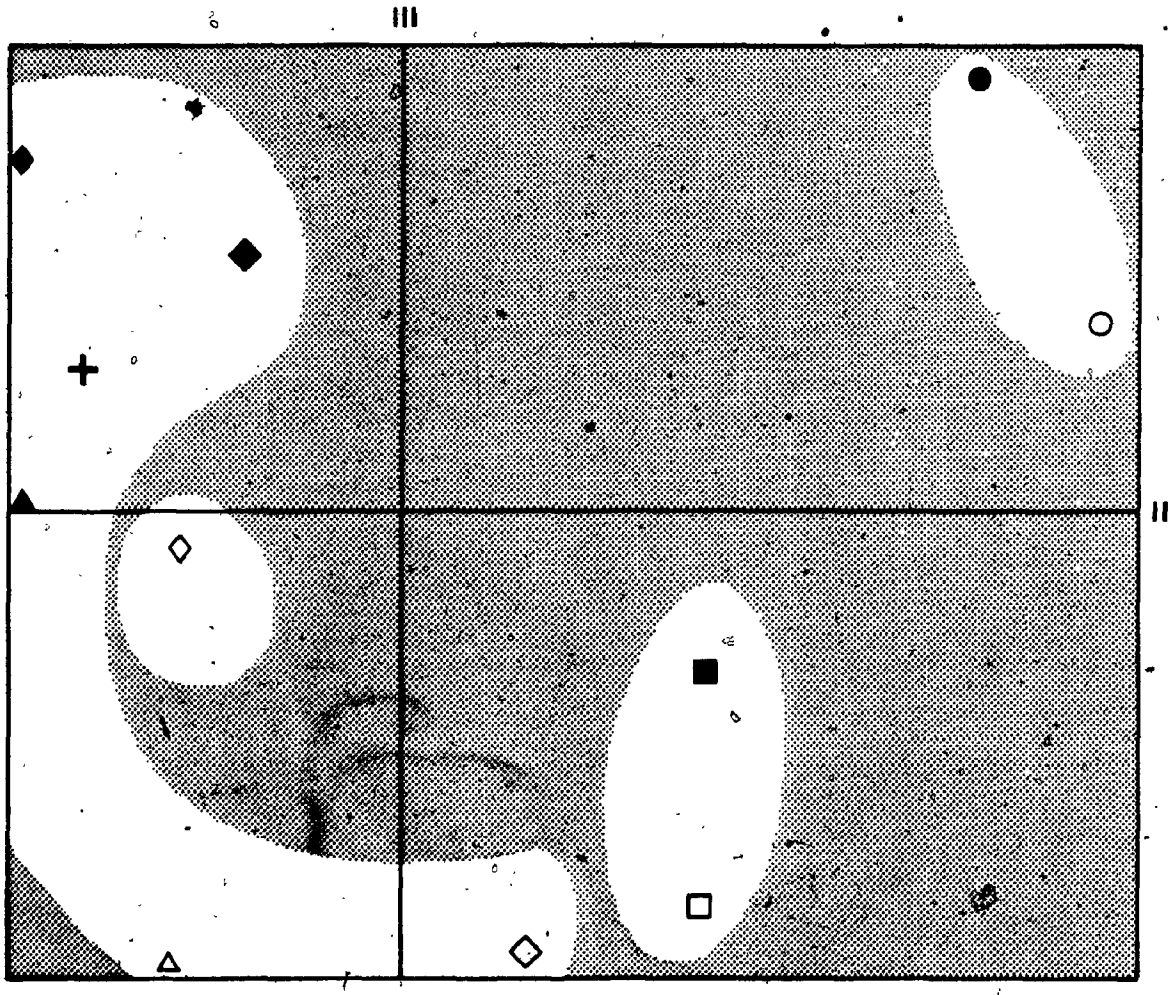


Figure 4.2 Eleven resemblance matrices in the space defined by the second and third principal components of their correlation matrix. Symbols as defined in Table 4.2. Unstippled areas indicate clusters existing at the level shown by the dotted line in Figure 4.1.

Table 4.1 Correlations between four resemblance matrices ($\underset{\sim}{E}$, $*\underset{\sim}{E}$, $\underset{\sim}{M}$, $\underset{\sim}{D}$) calculated from the flower and fruit data for 160 OTUs, and results of a Q-PCA of the four matrices based on this correlation matrix. See text for details, including explanation of the resemblance matrix abbreviations.

$*\underset{\sim}{E}$	0.88762		
$\underset{\sim}{M}$	0.64609	0.73396	
$\underset{\sim}{D}$	0.86507	0.92194	0.64958
	$\underset{\sim}{E}$	$*\underset{\sim}{E}$	$\underset{\sim}{M}$

Component	Eigenvalue	% trace	Component Scores			
			$\underset{\sim}{E}$	$*\underset{\sim}{E}$	$\underset{\sim}{M}$	$\underset{\sim}{D}$
I	3.173	86.8	1.065	0.857	0.755	0.858
II	0.288	7.9	0.012	-0.172	-0.294	0.415
III	0.192	5.3	-0.113	0.356	-0.230	-0.012

Table 4.2 Data sets and resemblance measures for 60 OTU subsample used in comparison of resemblance matrices. Note boldface abbreviations for resemblance matrices indicating data set and resemblance measure used. Symbols are those used to represent these matrices in Figure 4.1 and 4.2, as well as the phenograms derived from them, in Figure 6.16 and 6.17. See Table 3.1 and 3.3 for explanation of descriptor abbreviations.

DESCRIPTOR TYPES:

	Multistate	Continuous	
Flower, fruit	PROJ ANTH TCAL PUB1 PUB2	WFL LCAL LFR WFR	△ ▲ GRE ERE
Leaf	NUMSEC ANGSEC	X Y Z TEETH	○ ● GVE EVE
	◇ ◆ GMU EMU	□ ■ GCO ECO	All* ◇ ◆ GAL EAL

*including STYL and STAM.

Table 4.2 Cont.

RESEMBLANCE FUNCTIONS:

Solid symbols - Euclidean Distance matrices.

Open symbols - Generalized Distance matrices.

+ - Matrix of Information Radii among OTUs, based on all 11 flower and fruit descriptors (INF).

Table 4.3 Correlations among 11 resemblance matrices for 60 OTUs; results of Q-PCA of the 11 matrices.

GAL	0.623									
GRE	0.730	0.841								
GVE	0.453	0.659	0.428							
GCO	0.634	0.774	0.734	0.667						
GMU	0.673	0.771	0.775	0.526	0.515					
EAL	0.919	0.696	0.753	0.597	0.690	0.744				
ERE	0.911	0.678	0.801	0.456	0.664	0.740	0.960			
EVE	0.520	0.559	0.437	0.850	0.579	0.531	0.707	0.522		
ECO	0.705	0.661	0.663	0.615	0.790	0.493	0.746	0.686	0.675	
EMU	0.837	0.615	0.697	0.486	0.552	0.771	0.933	0.941	0.579	0.498
	INF	GAL	GRE	GVE	GCO	GMU	EAL	ERE	EVE	ECO

Component	Eigenvalue	% Trace	Range, Component Scores		
I	7.781	70.74	0.71	-	0.95
II	1.179	10.72	-0.34	-	0.61
III	0.736	6.70	-0.38	-	0.37
IV	0.616	5.60	-0.40	-	0.46
V	0.251	2.28	-0.27	-	0.27
VI	0.150	1.36	-0.22	-	0.18
VII	0.112	1.01	-0.18	-	0.20
VIII	0.091	0.83	-0.15	-	0.12
IX	0.071	0.65	-0.15	-	0.09
X	0.009	0.08	-0.08	-	0.04
XI	0.004	0.04	-0.05	-	0.02

CHAPTER FIVE

SEARCHING FOR GROUP STRUCTURE: CLUSTER ANALYSES

5.1 Introduction

As described in the preceding chapter, the flower and fruit data available for the entire 160 OTU sample was summarized as four 160 x 160 OTU resemblance matrices, E, *E, M and D. The present chapter is concerned with analyses based on these resemblances that have as their aim the detection of group structure in the sample.

The purpose of this search for group structure is to determine as objectively as possible the extent to which the data support classifying the sample into groups (clusters) of OTUs, and to compare these groups with the taxonomic affiliation and topodeme origin of the OTUs concerned (Table 5.1). In this way also the structure of the sample may be compared with existing reference classifications.

The methods used to detect groups are those of cluster analysis (Sneath & Sokal, 1973) or classification (Cormack, 1971; Orlóci, 1978). The following section describes the four sorting algorithms used here and in Chapter 6. All of these are methods that sequentially and agglomeratively find

groups which are hierarchical and non-overlapping (i.e. the SAHN category of methods, Sneath & Sokal, 1973). These have been employed because they comprise well-established, computationally efficient methods which produce polythetically defined hierarchical groups of OTUs. This methodology proceeds from multivariate estimates of pairwise resemblance between OTUs (Chapter 4) to the formation of groups by systematically adding increasingly dissimilar OTUs until the entire sample has been incorporated into a single group. The successive fusions and the resemblance levels at which they occur are summarized in Figures 5.1 through 5.12.

5.2 Sorting Algorithms

Since there was no way of knowing a priori the nature of the groups present in the sample, if any, a range of sorting algorithms was employed in order to help distinguish between data-dependent and method-dependent features of the results. At one extreme of the range considered was a method with which increasing group size resulted in decreased separation from other groups (single-linkage; a "space-contracting" method sensu Lance & Williams, 1967, cited by Orlóci, 1978). At the opposite extreme, two methods were used that are "space-dilating," with which increasing group size resulted in increased separation from other groups. The fourth method employed, average-linkage,

is intermediate between these two extremes, being "space-indifferent". With this method group separation is unaffected by group size. Each of these methods has its special advantages and disadvantages which are described in detail by Cormack (1971), Jardine and Sibson (1971), Sneath and Sokal (1973) and Orlóci (1978). Because all of these methods are well documented in the literature they will not be described further here. All four methods were implemented using an updated version of the program NCLAS (Podani, 1980). Preliminary analyses using these methods were also done using CLUSTAN-1C (Wishart, 1978).

Table 5.1 Entire sample of 160 OTUs arranged by sites (1 - 13; see Section 2.5, 2.6) and taxon (see Section 2.1), and indicating the symbols used in Figure 5.1 - 5.12 for each stratum. OTUs making up topodeme samples T1 through T7 are indicated by special symbols for each randomly sampled site (Fig. 2.1; Table 3.2). Deterministically sampled OTUs are indicated by diamonds (◆, ◇). Solid and open symbols are used to represent the 10- and 20- stamen morphotypes of C. crus-galli, respectively. Other taxa are indicated by the letters or symbols shown.

Table 5.1 Cont.

Site	Symbol	OTUs	Taxon	Features (1)
1	■ (T2)	101,106,108,110-112,114,115,118,120,121,123-126,129,130	10-stamen C. crus-galli	A-ivory (all OTUs)
	□ (T1)	104,105,116,117,119,122,127,131-133	20-stamen C. crus-galli	A-ivory (most OTUs)
2	● (T5)	226,401,416,421,424,428,430	10-stamen C. crus-galli	A-ivory, F-large
	◆	223	" " "	
	○ (T4)	204,207-209;402-404,406-412,414,417-420,422,423,425-427	20-stamen C. crus-galli	A-faint pink
	◇	224,752-754	" " "	anthesis earlier than that of T4)
	◈	771	C. punctata var. punctata	
	◈	772	C. punctata var. aurea	
3	▲ (T3)	301-316	10-stamen C. crus-galli	A-ivory
	◈	781	C. punctata var. punctata	
4	★ (T6)	603-608,610-622	10-stamen C. crus-galli	A-red, S-few
	★	696,697	intermediate form (2)	
	☆	601,602,691-695	C. ?grandis	A-red, F-large

Table 5.1 Cont.

Site	Symbol	OTUs	Taxon	Features (1)
5	P (T7)	501-504,507,508,511,512,514,516,517,519,520	C. punctata var. punctata	
	P (T7)	506,509,510,513,518	C. punctata var. aurea	
	M	769	C. succulenta	A-pink, C-densely toothed
6	b	801,803-806,809-813,815	C. sp. aff. C. bushii	C-densely toothed
	◆	802,807,808,814	10-stamen C. crus-galli	A-ivory
7	◆	791-793	10-stamen C. crus-galli	A-red, projection of connective (PROJ) present
8	◆	703	10-stamen C. crus-galli	A-ivory
	◆	702,704	" " "	A-red
9	◆	711,712	10-stamen C. crus-galli	A-ivory
	◇	713,714	20-stamen C. crus-galli	A-pink, faint pink
10	◆	731	10-stamen C. crus-galli	A-ivory
11	◇	751	20-stamen C. crus-galli	A-pink

Table 5.1 Cont.

Site	Symbol	OTUs	Taxon	Features (1)
12	◆	721-722	10-stamen C. crus-galli	A-ivory
13	M	761-762	C. macracantha	A-faint pink, C-densely toothed

(1) A: color of undehisced anthers (ANTH).
S: number of styles (STYL).
F: fruit size (LFR, WFR).
C: calyx lobe margination (TCAL).

(2) OTUs intermediate between sympatric 10-stamen C. crus-galli and C. ?grandis.

5.3 Results

Because of the large sample size, the fusions of similar OTUs are not represented in the phenograms. The behavior of individual OTUs is shown in Figures 5.1 through 5.12 only where because of the manner in which fusions occur (sequence, resemblance level), either distinct groups are not formed, or if they are, they contain fewer than five or six OTUs.

One feature that is common to the twelve phenograms produced here (Fig. 5.1 - 5.12) is the coherence of the individual topodeme samples. With few exceptions most of the OTUs making up a single topodeme sample clustered together before joining with OTUs from other sites. Noise is introduced into this pattern, however, by the presence in the 160 OTU sample of additional, deterministically sampled OTUs (Table 5.1).

The results obtained with the three sorting algorithms are presented in the following sections. For the most part they conform to one of three patterns of hierarchical relationships among the individual clusters of OTUs. In the first (pattern A) the OTUs are resolved into only three principal groups (Fig. 5.5, 5.8, 5.9, 5.11):

(punctata, ?grandis)
((bushii, 10-stamen crus-galli)(20-stamen crus-galli))

In the second pattern (B) the sample is sorted into two groups according to stamen number, as follows (Fig. 5.10, 5.12):

(bushii, 10-stamen crus-galli)
((punctata, ?grandis)(20-stamen crus-galli))

In comparison with A and B, pattern C is somewhat more structured (Fig. 5.1, 5.2):

(punctata, ?grandis)
(bushii (T6 (10-stamen crus-galli)(20-stamen crus-galli)))

In describing these patterns, the three OTUs belonging to section Macracanthae (761, 762, 769) have not been considered, but their behavior will be discussed in Section 5.4. Intermediate patterns found in Figures 5.3, 5.4, 5.6 and 5.7 are discussed below under the sorting algorithm concerned.

5.3.1 Single-linkage

As is commonly the case, single-linkage clustering produced the most equivocal group structure. This was particularly true with those resemblance matrices whose calculation minimizes the effects of descriptor correlations (generalized distances, M, and information radii, D). The phenogram produced from M does not fit any one of patterns A-C because of the way in which the sample of C. sp. aff. C. bushii is set among the OTUs of C. punctata (Fig. 5.3).

The hierarchical relationship among the clusters produced from D (Fig. 5.4), however, is intermediate between those of patterns A and C in the way in which at least the bushii OTUs are distinguished from the remainder of that portion of the sample. The more structured pattern C only emerged in the phenograms obtained from the euclidean distance matrices (E, *E; Fig. 5.1, 5.2).

5.3.2 Average-linkage

The clusters of OTUs produced by average-linkage clustering were better defined than those obtained above, in terms of their greater contrast between the resemblance level at which clusters formed, and that at which they fused with other clusters. The phenogram obtained from M departed from pattern A by the way in which intermediate OTUs of 10- and 20-stamen C. crus-galli (branches G, I; Fig. 5.7) bracketed the sample of C. sp. aff. C. bushii (branch H). The remaining average-linkage phenograms exhibit pattern A, or nearly so. The phenogram produced from *E (Fig. 5.6) resembles Figure 5.4 and also approaches pattern C in the way in which the bushii OTUs are distinguished.

5.3.3 Minimum Variance

Of the two space-dilating sorting methods described in Section 5.2 above, only the minimum variance method has been

employed with the 160 OTU sample. Complete-linkage, however, has been used in the analyses described in the following chapter (Section 6.3). Here, minimum variance clustering produced pattern A with both $\underset{\sim}{E}$ and $\underset{\sim}{M}$ (Fig. 5.9, 5.11). With the euclidean distances based on standardized data ($^{*}\underset{\sim}{E}$; Fig. 5.10) and with the information radii ($\underset{\sim}{D}$; Fig. 5.12), however, use of this algorithm apparently generated groups of OTUs on the basis of stamen number (pattern B).

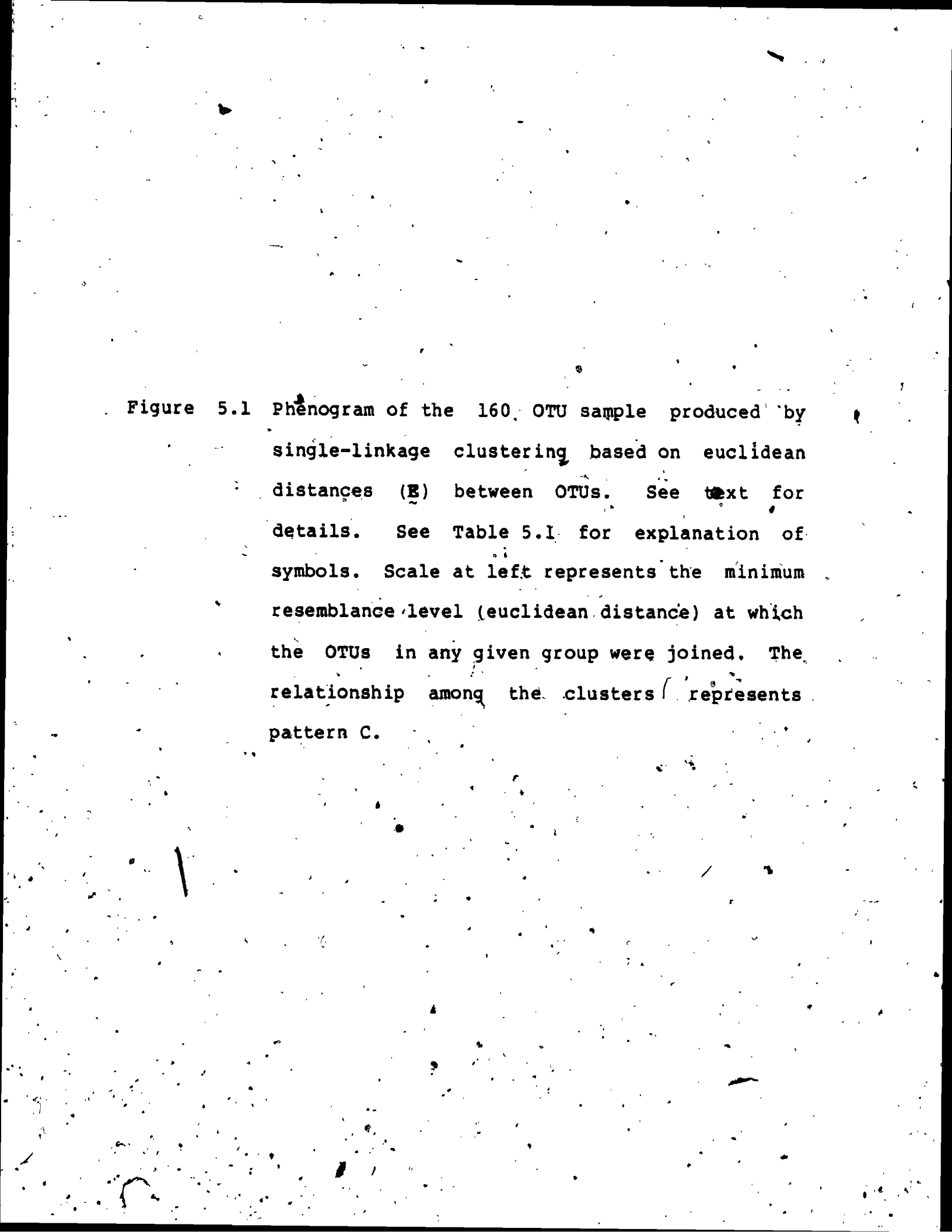


Figure 5.1 Phenogram of the 160 OTU sample produced by single-linkage clustering based on euclidean distances (E) between OTUs. See text for details. See Table 5.1 for explanation of symbols. Scale at left represents the minimum resemblance level (euclidean distance) at which the OTUs in any given group were joined. The relationship among the clusters represents pattern C.

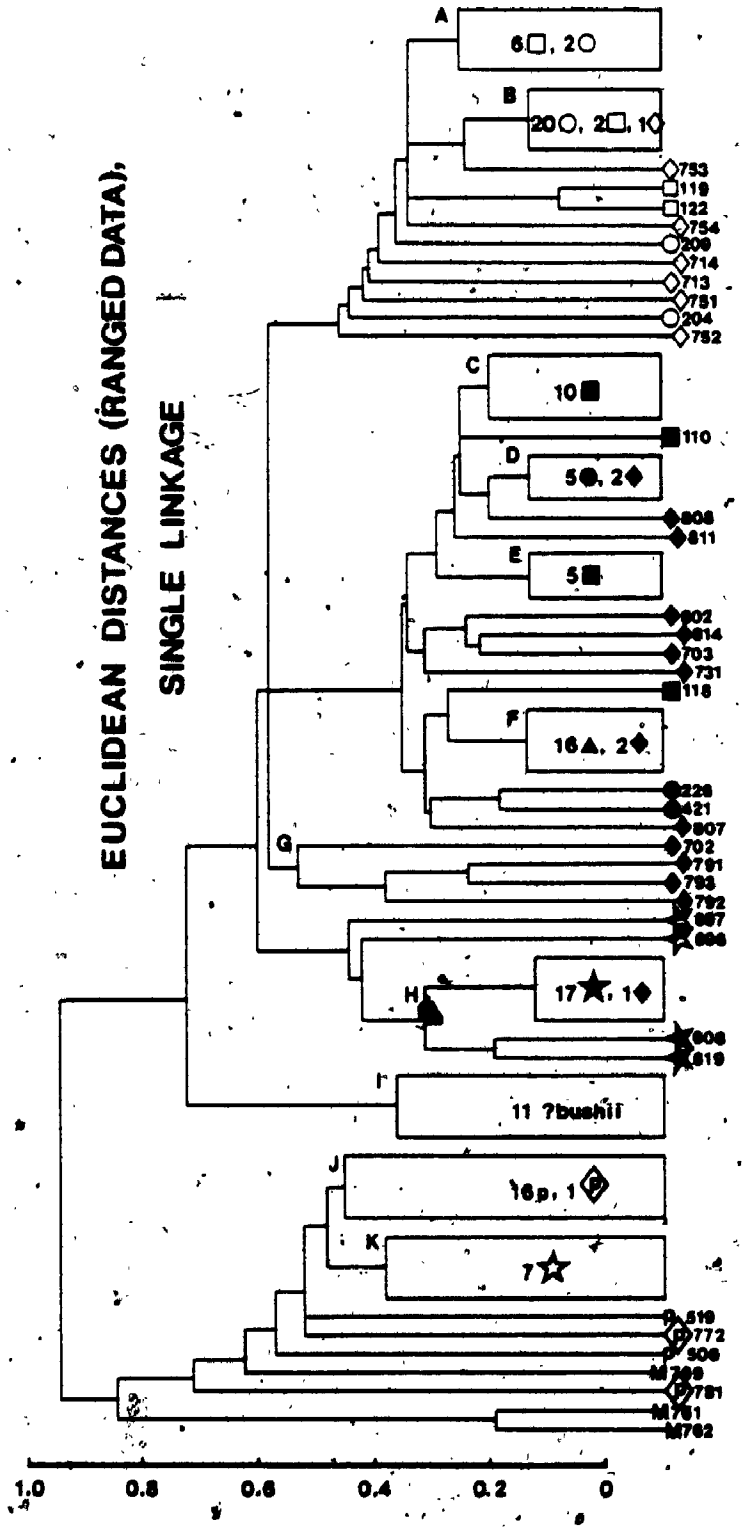

EUCLIDEAN DISTANCES (RANGED DATA),
SINGLE LINKAGE
A
6□, 2○
B
20○, 2□, 1◇
753
119
122
754
209
714
713
751
204
752
C
10■
110
D
5●, 2◆
808
811
E
5■
802
814
703
731
116
F
16▲, 2◆
228
421
807
702
791
793
792
797
806
H
17★, 1◆
808
819
I
11 ?bushii
J
16p, 1
K
7☆
519
772
506
M760
781
M761
M762
1.0
0.8
0.6
0.4
0.2
0

Figure 5.2 Phenogram of the 160 OTU sample produced by single-linkage clustering based on euclidean distances (*E) between OTUs. See text for details. See Table 5.1 for explanation of symbols. Scale at left represents the minimum resemblance level (euclidean distance) at which the OTUs in any given group were joined. The relationship among the clusters represents pattern C.

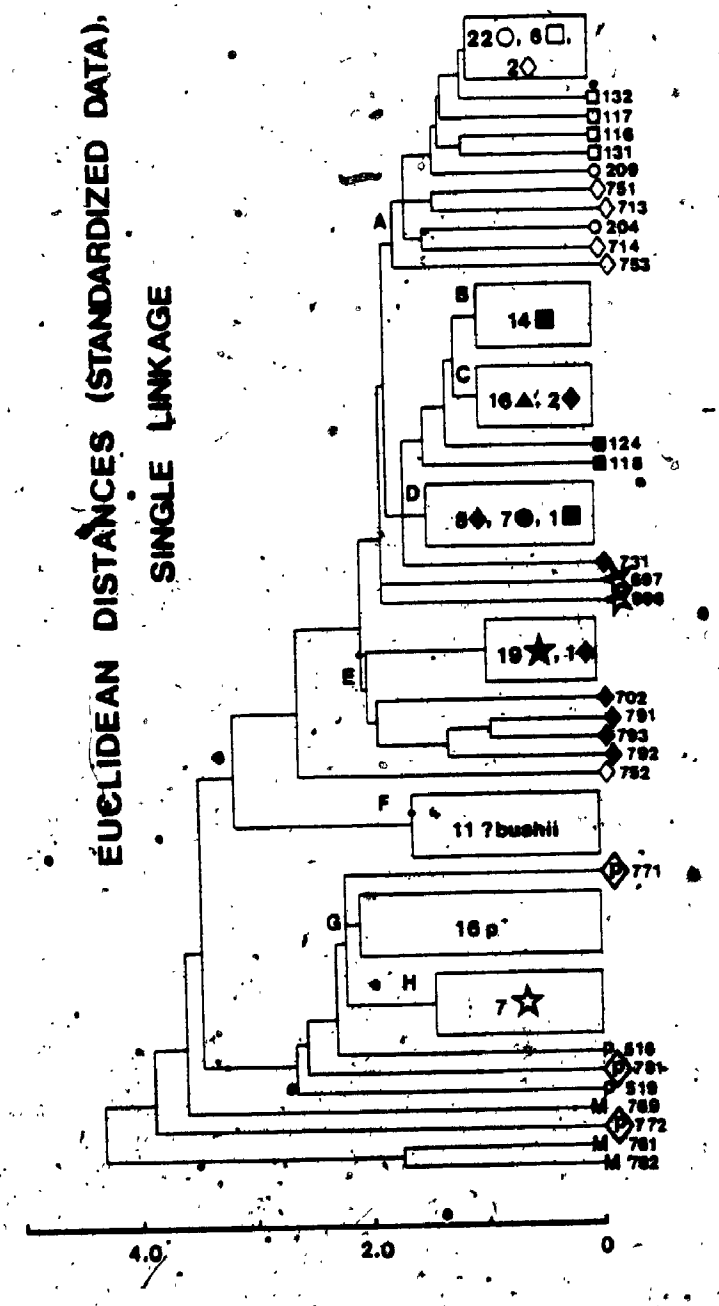

EUCLIDEAN DISTANCES (STANDARDIZED DATA),
SINGLE LINKAGE
22○, 6□,
2◇
132
117
116
131
209
751
713
204
714
753
A
B
14■
C
16▲, 2◆
124
118
D
8◆, 7●, 1■
731
687
696
19★, 1◆
E
702
791
793
792
752
F
11 ?bushii
771
G
16 p
H
7☆
516
781
519
769
772
781
782
4.0
2.0
0

Figure 5.3 Phenogram of the 160 OTU sample produced by single-linkage clustering based on generalized distances (M) between OTUs. See text for details. See Table 5.1 for explanation of symbols. Scale at left represents the minimum resemblance level (generalized distance) at which the OTUs in any given group were joined. The relationship among the clusters does not correspond to any one of patterns A - C.

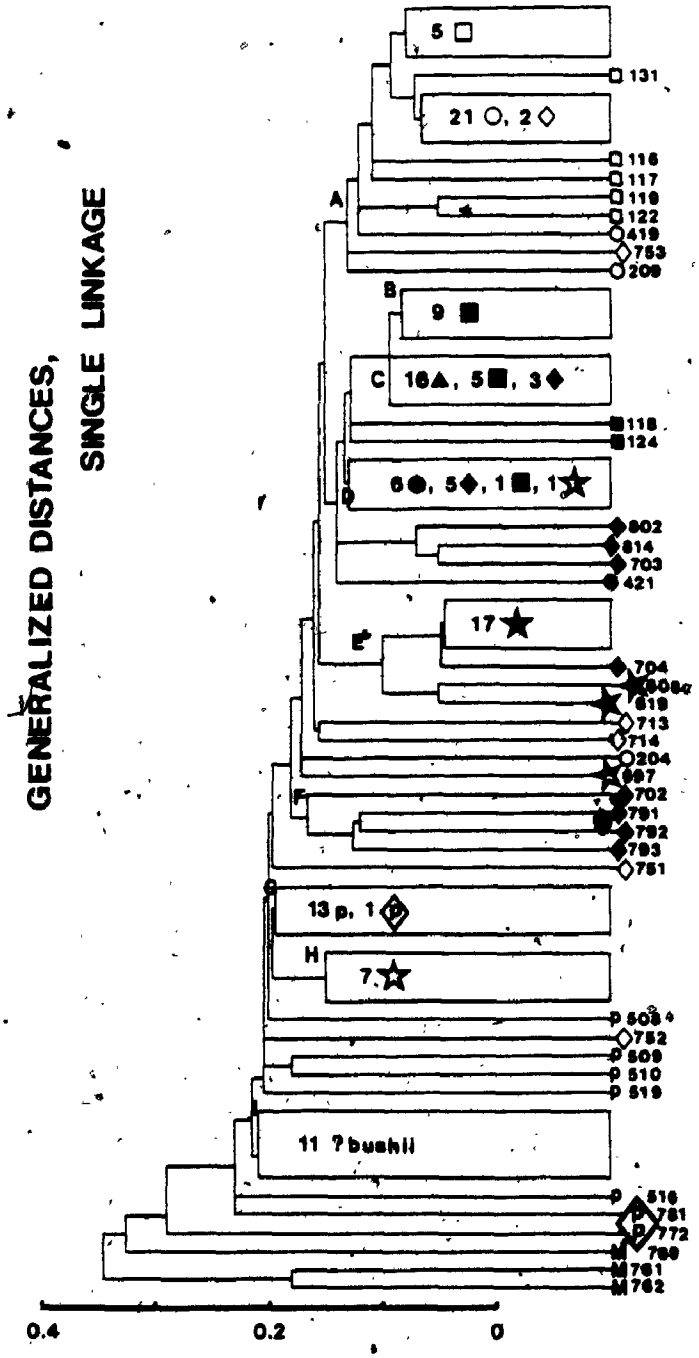

GENERALIZED DISTANCES, SINGLE LINKAGE
5 □
131
21 ○, 2 ◇
116
117
119
122
419
753
209
A
B
9 ■
C
16▲, 5■, 3◆
118
124
6●, 5◆, 1■, 1★
802
814
703
421
17★
704
819
713
714
204
F
702
791
792
793
751
13 p, 1
H
7 ☆
752
509
510
519
11 ? bushii
516
781
772
760
761
762
0.4
0.2
0

Figure 5.4 Phenogram of the 160 OTU sample produced by single-linkage clustering based on information radii (D) between OTUs. See text for details. See Table 5.1 for explanation of symbols. Scale at left represents the minimum resemblance level (information radius) at which the OTUs in any given group were joined. The relationship among the clusters does not correspond to any one of the patterns A - C.

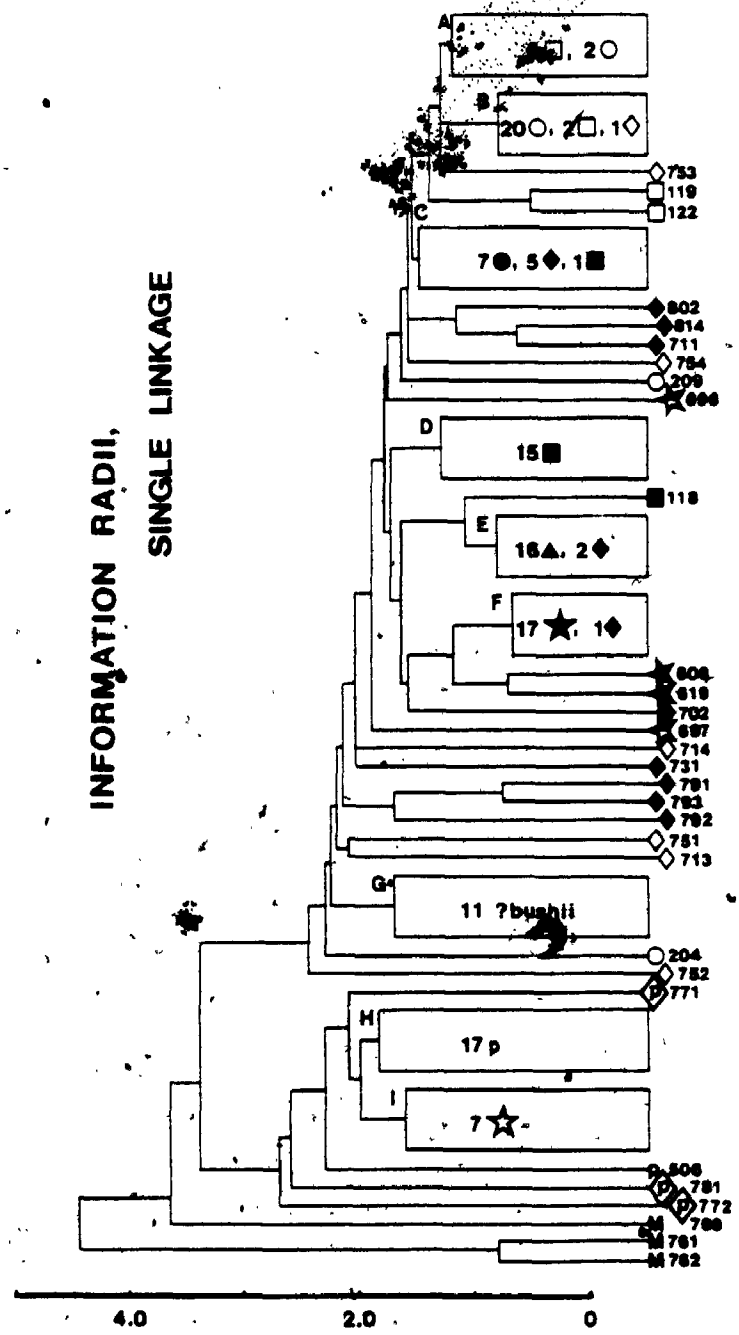

INFORMATION RADII,
SINGLE LINKAGE
A
2○
B
20○, 2□, 1◇
753
119
122
C
7●, 5◆, 1■
802
814
711
754
209
696
D
15■
118
E
16▲, 2◆
F
17★, 1◆
806
618
702
697
714
731
781
783
782
751
713
G
11 ?bushii
204
752
771
H
17 p
I
7☆
506
781
772
768
781
782
4.0
2.0
0

Figure 5.5 Phenogram of the 160 OTU sample produced by average-linkage clustering based on euclidean distances (E) between OTUs. See text for details. See Table 5.1 for explanation of symbols. Scale at left represents the minimum resemblance level (euclidean distance) at which the OTUs in any given group were joined. The relationship among the clusters represents pattern A.

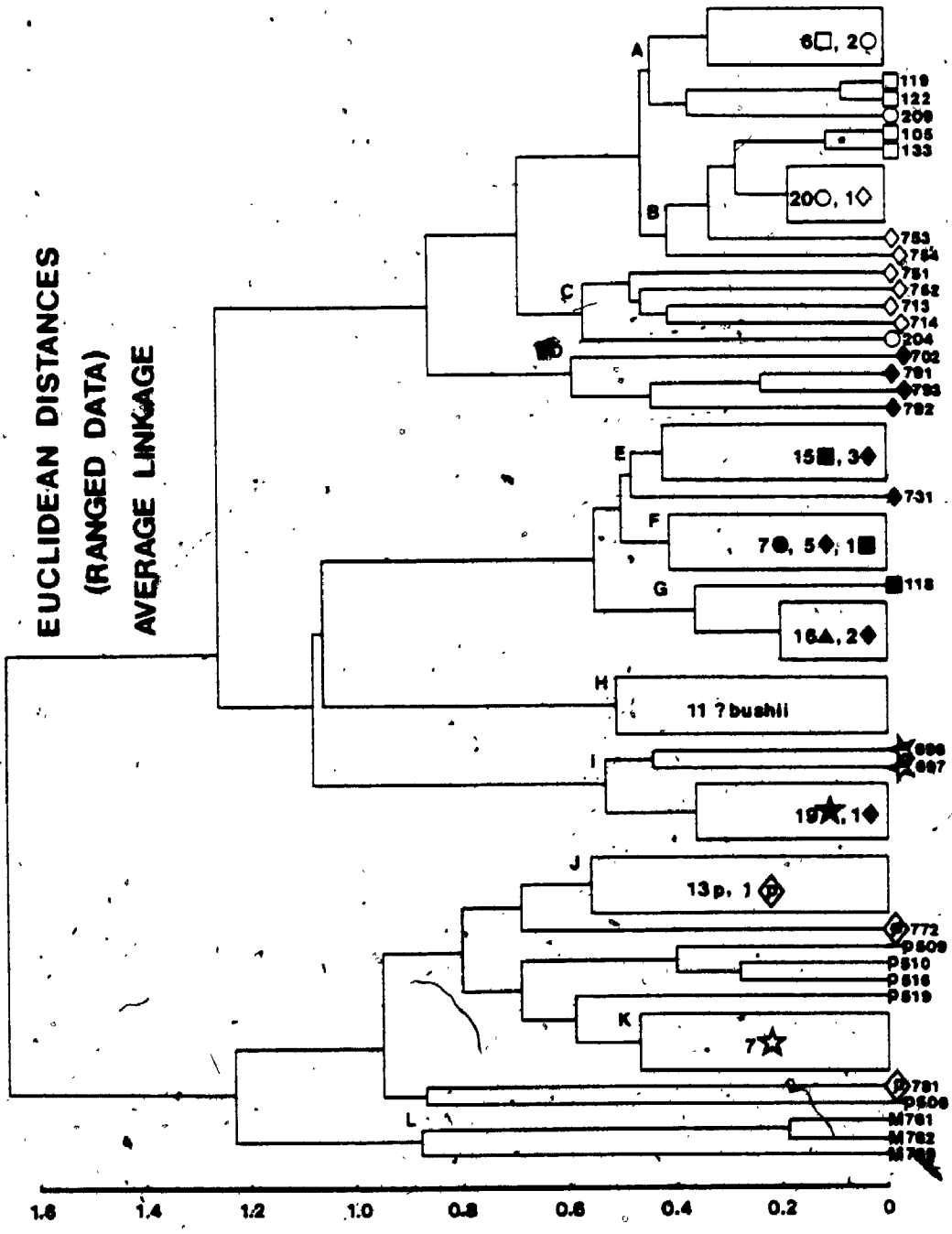
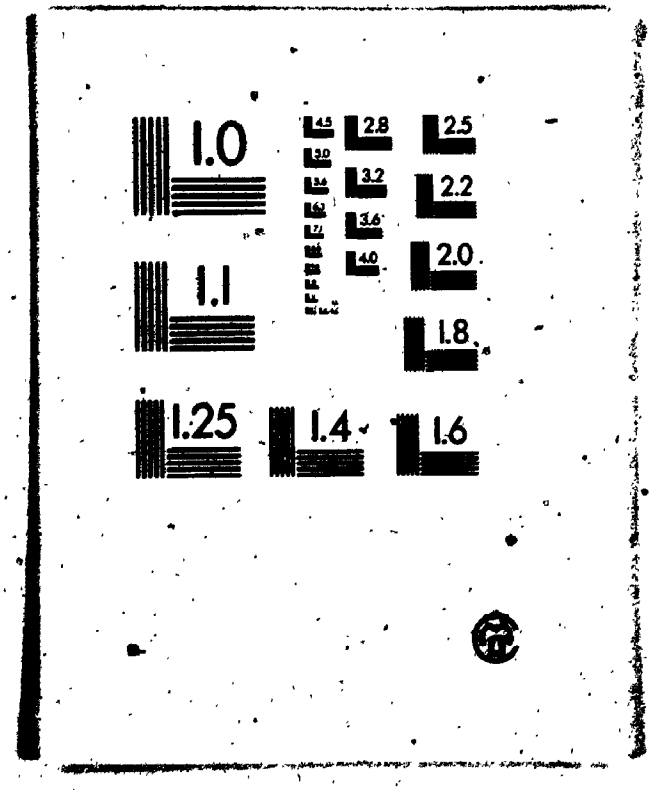

EUCLIDEAN DISTANCES
(RANGED DATA)
AVERAGE LINKAGE
A
6□, 2○
119
122
209
105
133
B
20○, 1◇
753
754
C
751
752
713
714
204
D
702
791
793
792
E
15■, 3◆
731
F
7●, 5◆, 1■
G
118
16▲, 2◆
H
11 ? bushii
I
696
697
19★, 1◆
J
13p, 1
772
p509
p510
p516
p519
K
7☆
781
p508
M781
M782
M783
L
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0

Figure 5.6 Phenogram of the 160 OTU sample produced by average-linkage clustering based on euclidean distances (*E) between OTUs. See text for details. See Table 5.1 for explanation of symbols. Scale at left represents the minimum resemblance level (euclidean distance) at which the OTUs in any given group were joined. The relationship among the clusters does not correspond to any one of the patterns A - C.

3



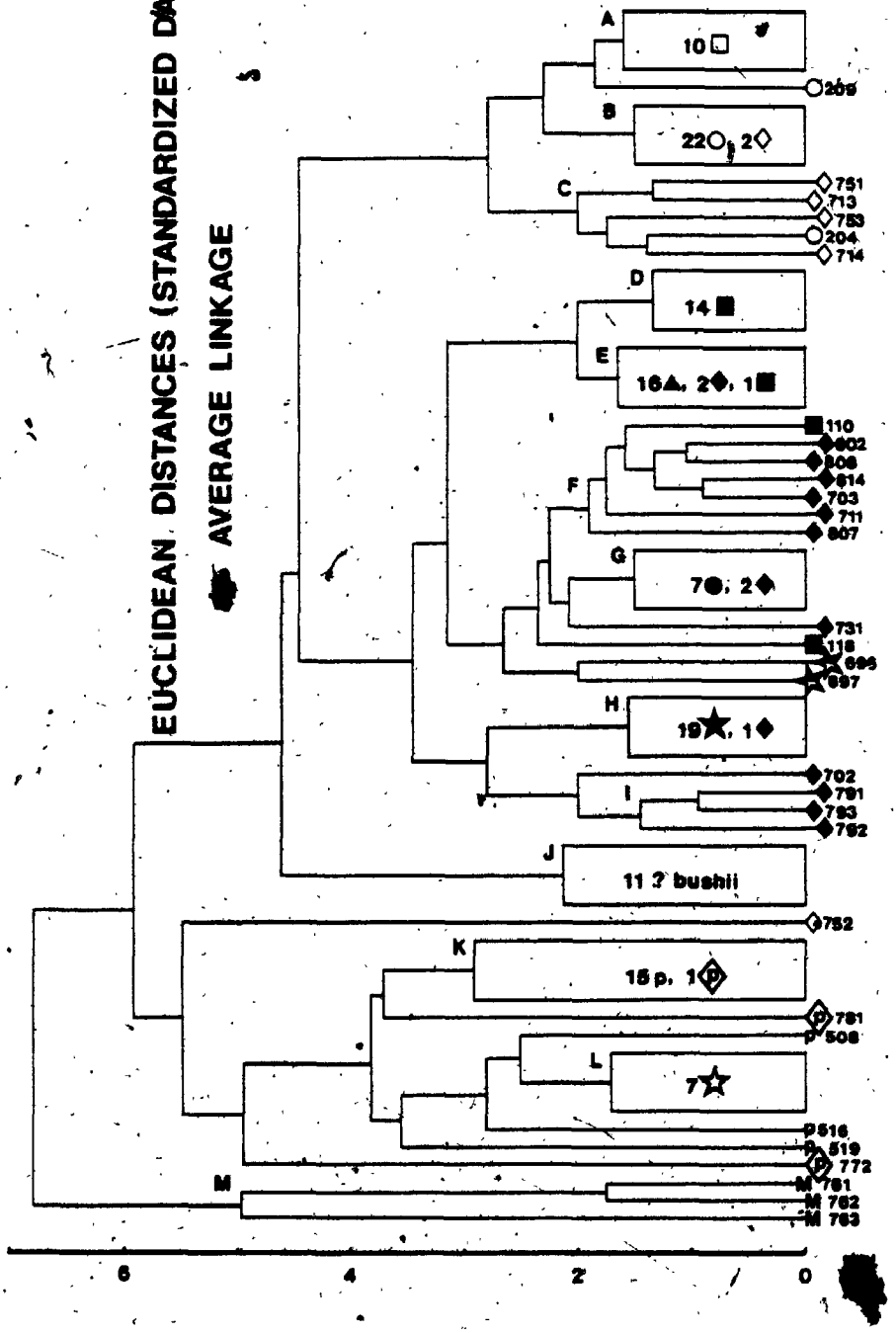

EUCLIDEAN DISTANCES (STANDARDIZED DATA),
AVERAGE LINKAGE
A
10
B
22, 2
C
751
713
753
204
714
D
14
E
16, 2, 1
110
802
806
814
703
711
807
F
G
7, 2
731
118
695
697
H
19, 1
702
791
793
792
I
J
11 ? bushii
752
K
15 p, 1
791
p 508
L
7
p 516
p 519
772
M
M 781
M 782
M 783
6
4
2
0

Figure 5.7 Phenogram of the 160 OTU sample produced by average-linkage clustering based on generalized distances (M) between OTUs. See text for details. See Table 5.1 for explanation of symbols. Scale at left represents the minimum resemblance level (generalized distance) at which the OTUs in any given group were joined. The relationship among the clusters does not correspond to any one of patterns A - G.

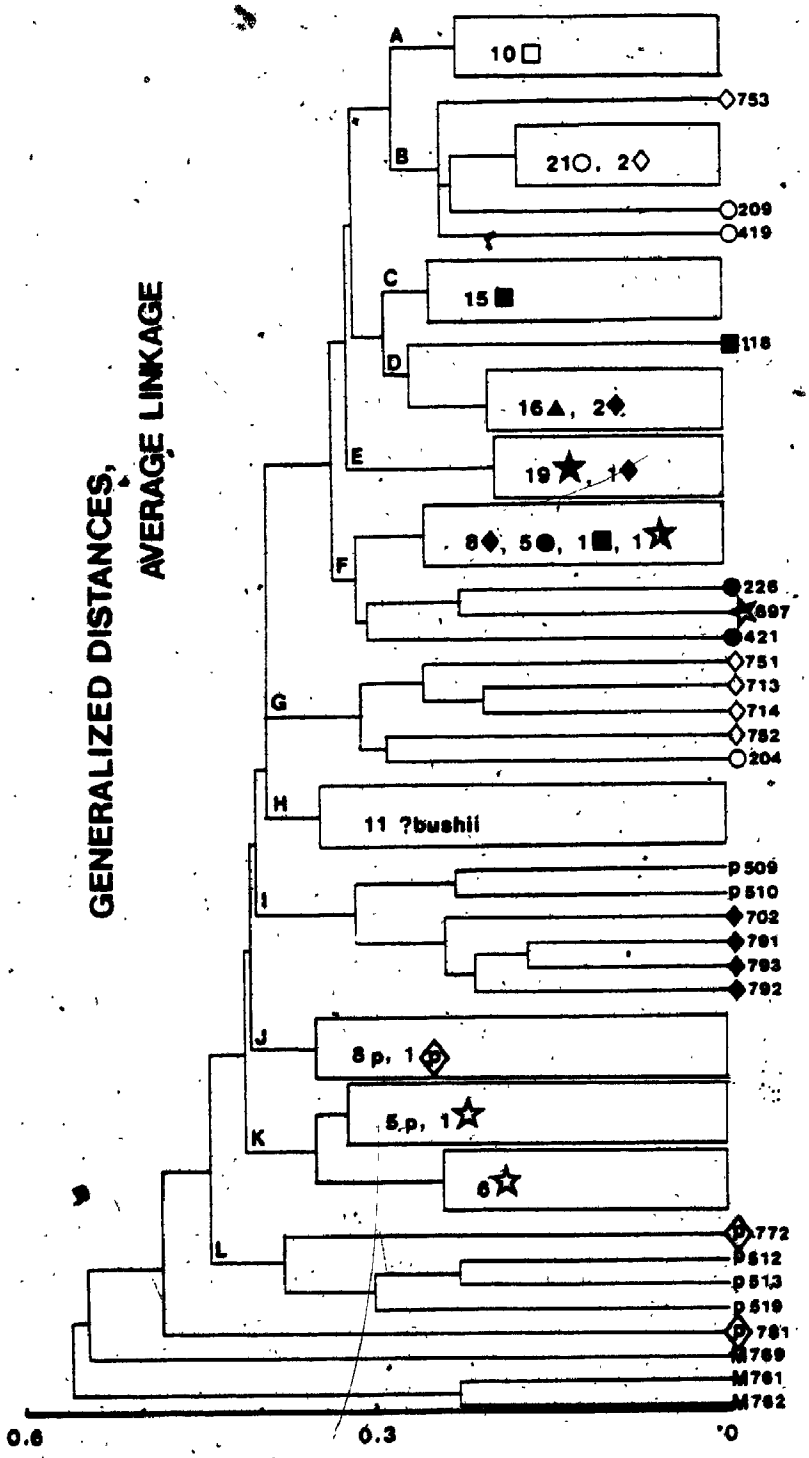

GENERALIZED DISTANCES, AVERAGE LINKAGE
A
10 □
753
B
21○, 2◇
209
419
C
15 ■
D
118
16▲, 2◆
E
19★, 1◆
F
8◆, 5●, 1■, 1★
226
897
421
G
751
713
714
782
204
H
11 ?bushii
I
p509
p510
702
791
793
792
J
8 p, 1
K
5 p, 1☆
6☆
L
772
p512
p513
p519
791
M769
M781
M782
0.6
0.3
0

Figure 5.8 Phenogram of the 160 OTU sample produced by average-linkage clustering based on information radii (D) between OTUs. See text for details. See Table 5.1 for explanation of symbols. Scale at left represents the minimum resemblance level (information radius) at which the OTUs in any given group were joined. The relationship among the clusters represents pattern A.

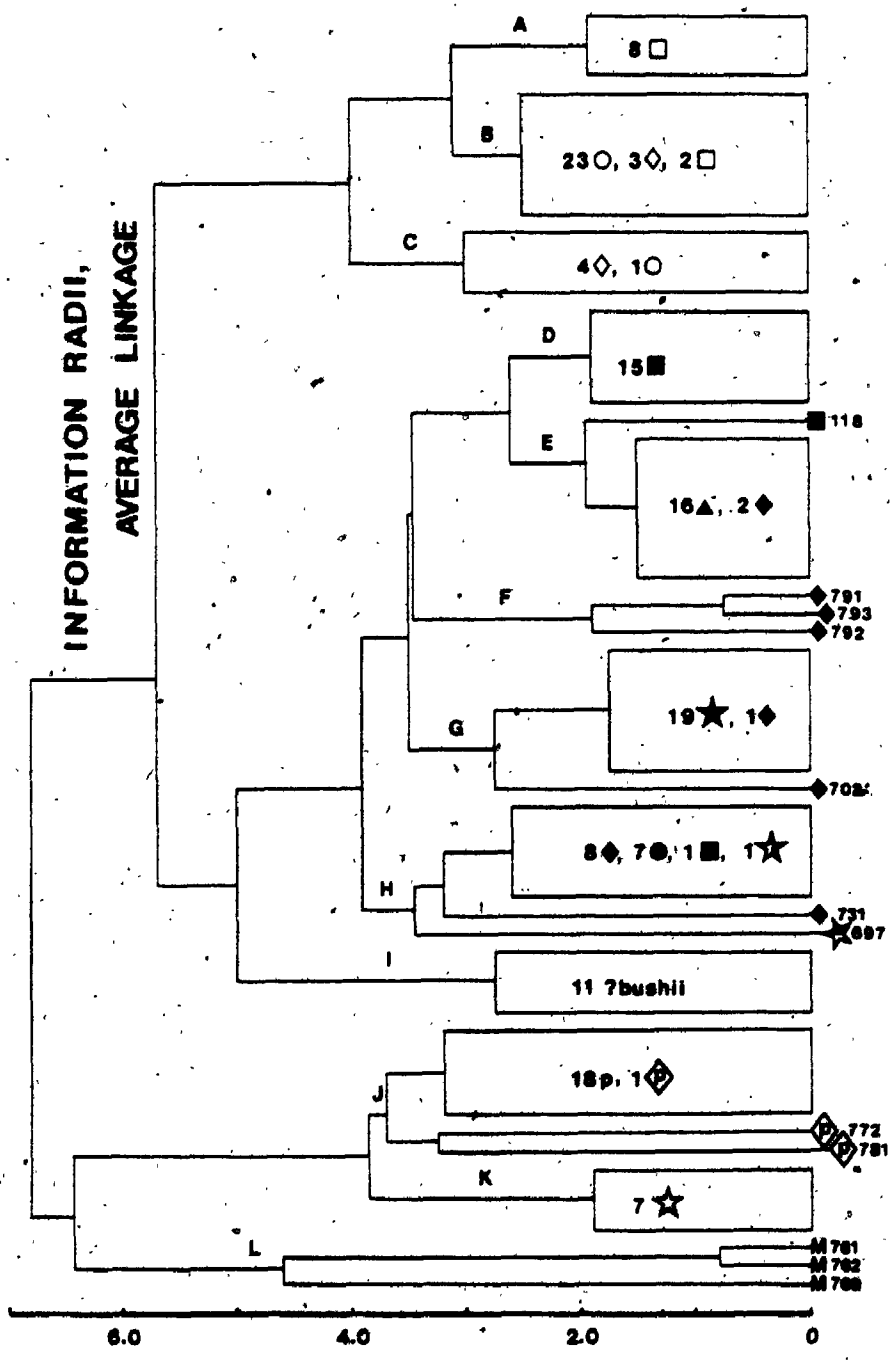

INFORMATION RADII, AVERAGE LINKAGE
A
8 □
B
23 ○, 3 ◇, 2 □
C
4 ◇, 1 ○
D
15 ■
118
E
16 ▲, 2 ◆
F
791
793
792
19 ★, 1 ◆
G
8 ◆, 7 ●, 1 ■, 1 ☆
H
731
697
I
11 ?bushii
18 p, 1 ⊗
J
772
781
K
7 ☆
L
M 781
M 782
M 788
6.0
4.0
2.0
0

Figure 5.9 Phenogram of the 160 OTU sample produced by minimum variance clustering based on euclidean distances (E) between OTUs. See text for details. See Table 5.1 for explanation of symbols. Scale at left represents the within-group sum of squares for each new group of OTUs. The relationship among the clusters represents pattern A.

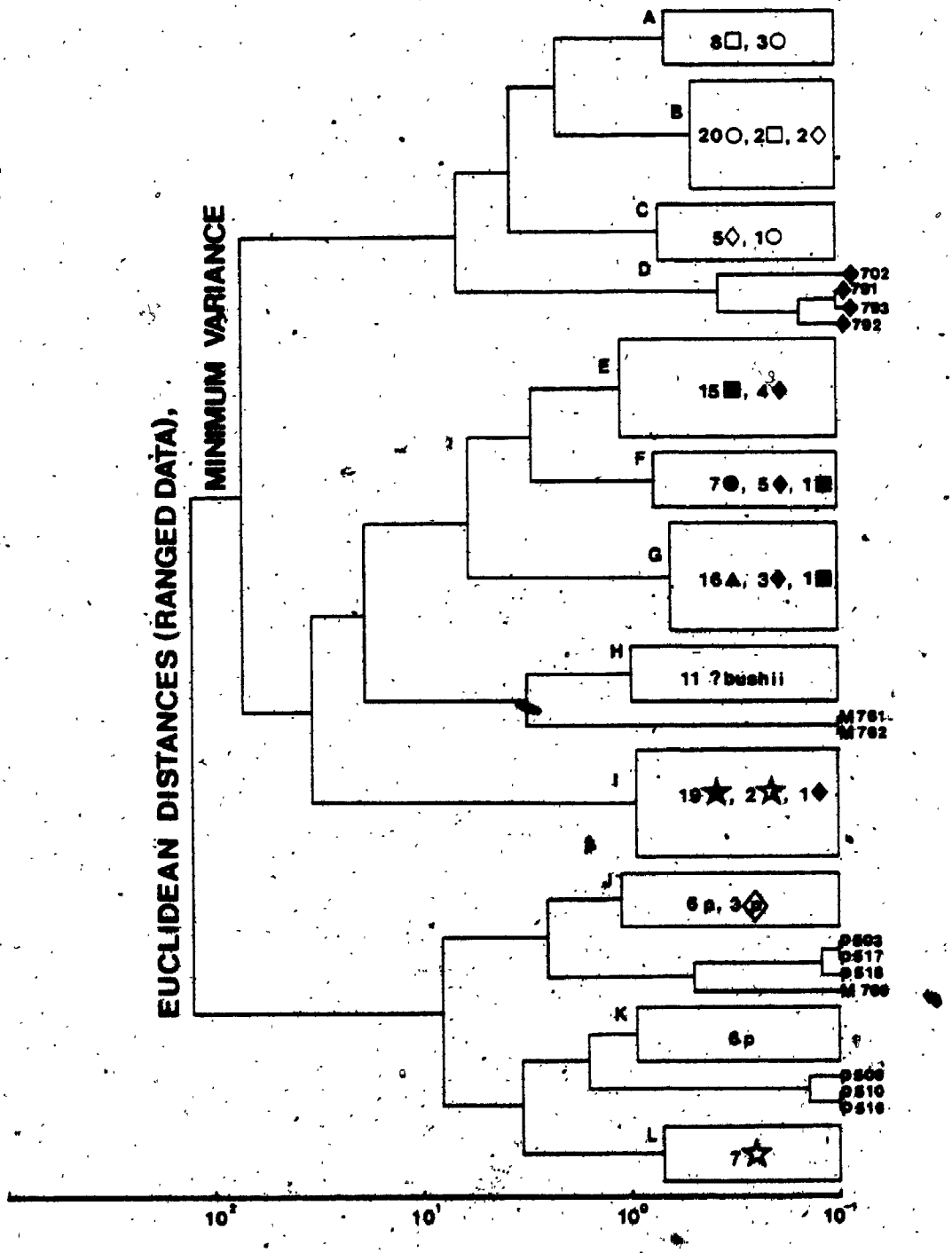

EUCLIDEAN DISTANCES (RANGED DATA), MINIMUM VARIANCE
A
8□, 3○
B
20○, 2□, 2◇
C
5◇, 1○
D
E
15■, 4◆
F
7●, 5◆, 1■
G
16▲, 3◆, 1■
H
11 ?bushii
I
19★, 2☆, 1◆
J
6 p, 3
K
6p
L
7☆
10²
10¹
10⁰
10⁻¹

Figure 5.10 Phenogram of the 160 OTU sample produced by minimum variance clustering based on euclidean distances (*E) between OTUs. See text for details. See Table 5.1 for explanation of symbols. Scale at left represents the within-group sum of squares for each new group of OTUs. The relationship among the clusters represents pattern B.

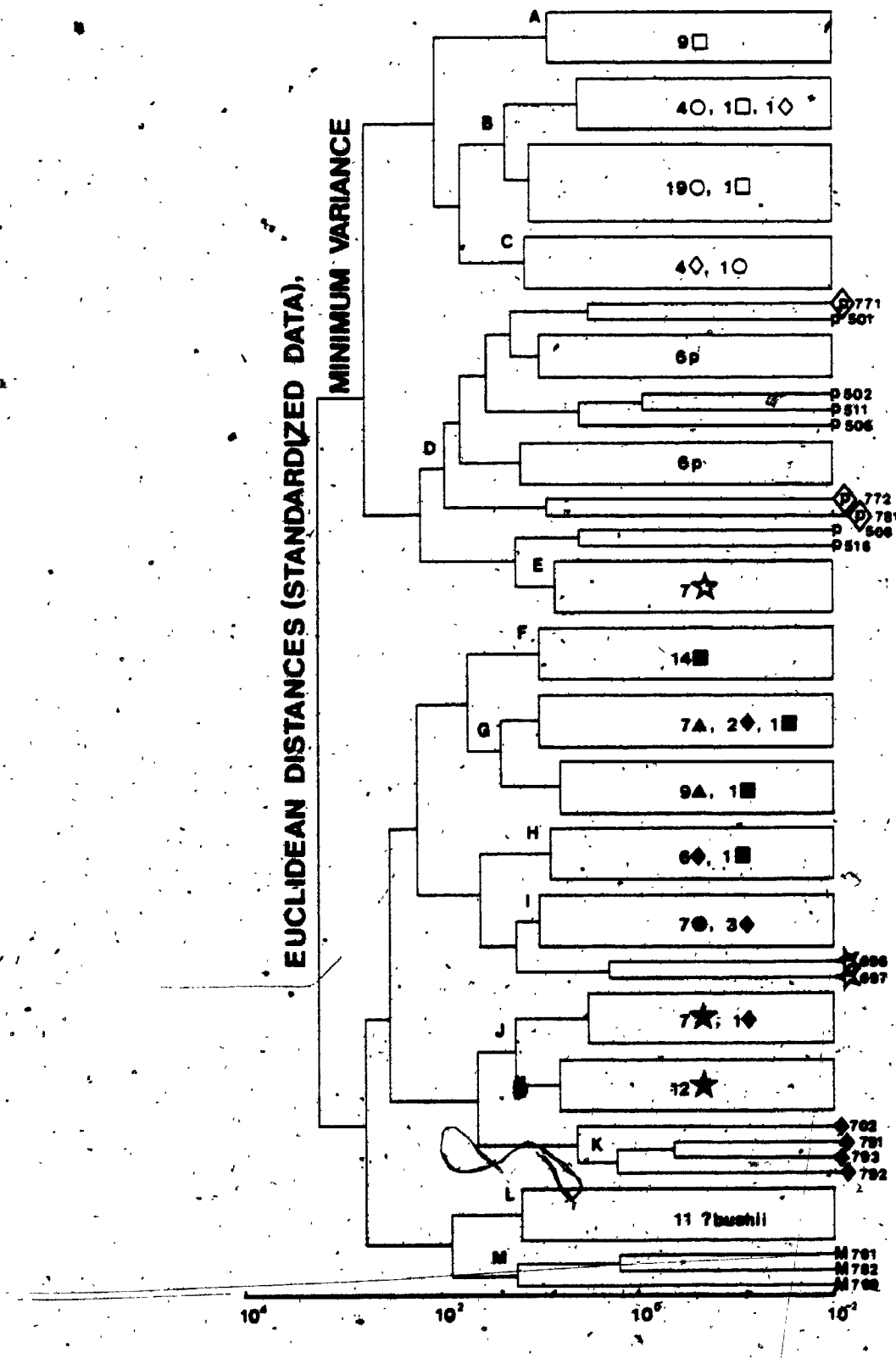

EUCLIDEAN DISTANCES (STANDARDIZED DATA), MINIMUM VARIANCE
A
9□
B
4○, 1□, 1◇
19○, 1□
C
4◇, 1○
6p
6p
D
6p
7☆
E
14■
F
7▲, 2◆, 1■
G
9▲, 1■
6◆, 1■
H
I
7●, 3◆
7★, 1◆
J
12★
K
11 ?bushii
L
M
10^{4}
10^{2}
10^{0}
10^{-2}

Figure 5.11 Phenogram of the 160 OTU sample produced by minimum variance clustering based on generalized distances ($\tilde{M}$) between OTUs. See text for details. See Table 5.1 for explanation of symbols. Scale at left represents the within-group sum of squares for each new group of OTUs. The relationship among the clusters represents pattern A.

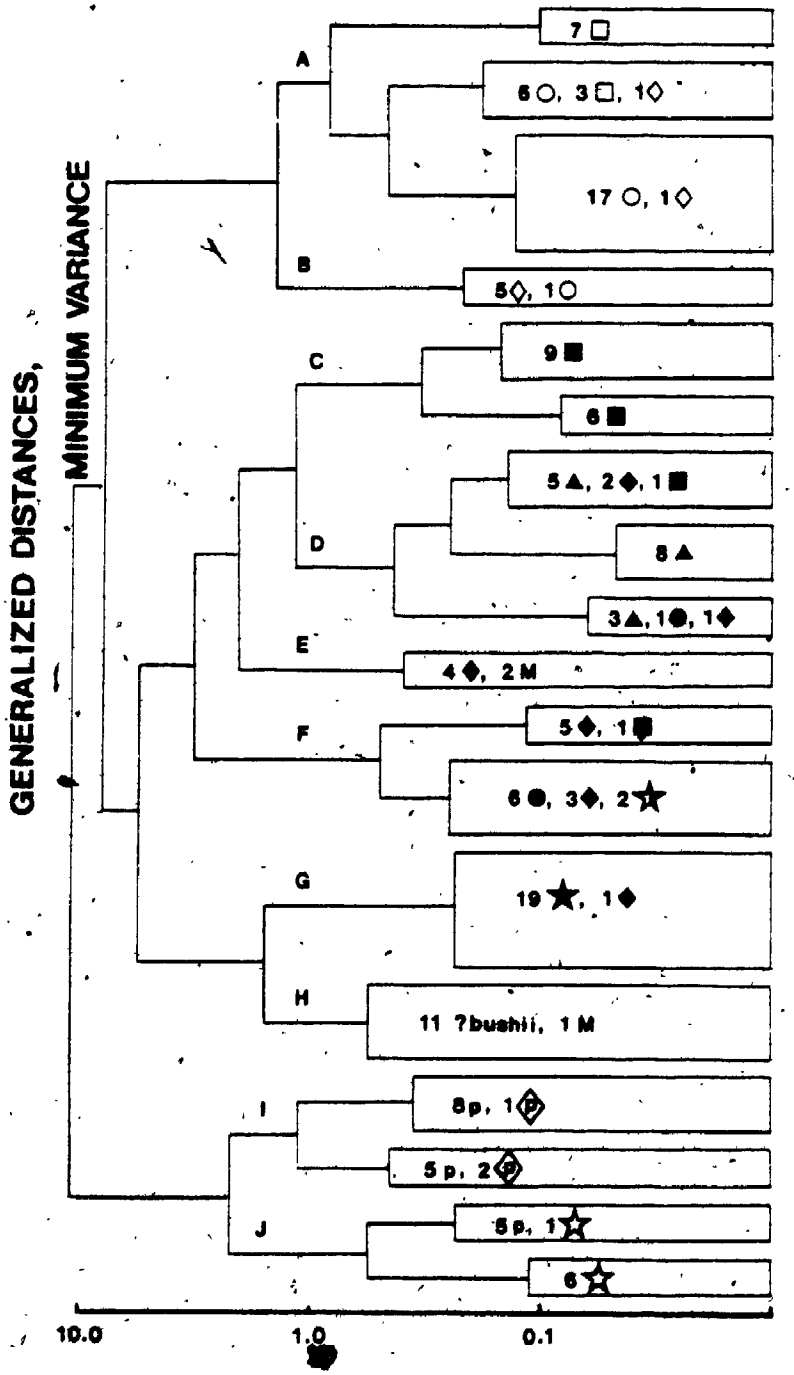
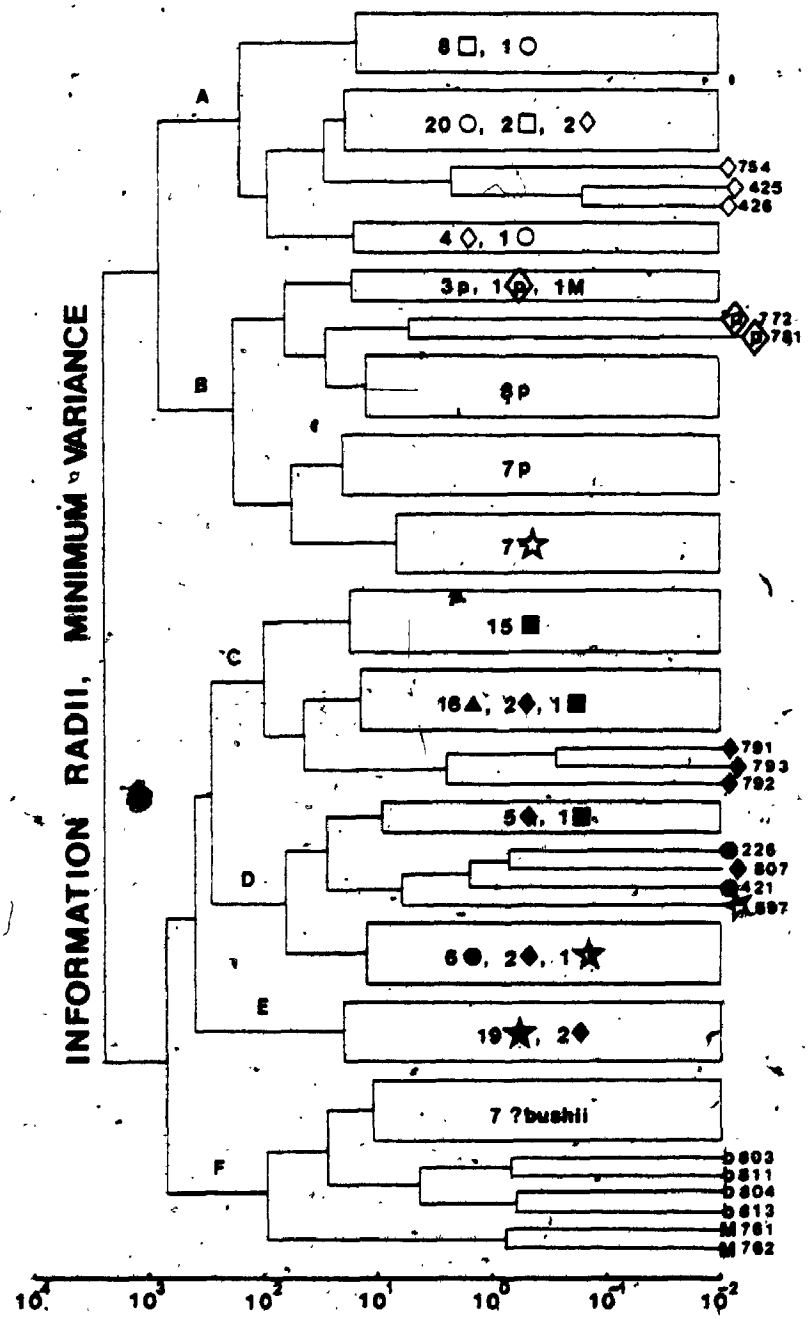

GENERALIZED DISTANCES, MINIMUM VARIANCE
A
B
C
D
E
F
G
H
I
J
7 □
6 ○, 3 □, 1 ◇
17 ○, 1 ◇
5 ◇, 1 ○
9 ■
6 ■
5 ▲, 2 ◆, 1 ■
8 ▲
3 ▲, 1 ●, 1 ◆
4 ◆, 2 M
5 ◆, 1 ■
6 ●, 3 ◆, 2 ☆
19 ★, 1 ◆
11 ?bushii, 1 M
8p, 1 ◈
5p, 2 ◈
5p, 1 ☆
6 ☆
10.0
1.0
0.1

Figure 5.12 Phenogram of the 160 OTU sample produced by minimum variance clustering based on the information radii (D) between OTUs. See text for for details. See Table 5.1 for explanation of symbols. Scale at left represents the within-group sum of squares for each new group of OTUs. The relationship among the clusters represents pattern B.


INFORMATION RADII, MINIMUM VARIANCE
8 □, 1 ○
20 ○, 2 □, 2 ◇
754
425
426
4 ◇, 1 ○
3p, 1 ◇, 1M
772
781
8p
7p
7☆
15 ■
16▲, 2◆, 1■
791
793
792
5◆, 1■
226
807
421
597
6●, 2◆, 1☆
19★, 2◆
7 ?bushii
b803
b811
b804
b813
M781
M782
A
B
C
D
E
F
10⁴
10³
10²
10¹
10⁰
10⁻¹
10⁻²

5.4 Discussion

The most consistent result in all the cluster analyses reported here was the recovery of all or most of the individual topodeme samples as clusters of OTUs. In addition, although twelve different combinations of resemblance coefficient and sorting algorithms were used, eight of these produced only three patterns (A-C) of hierarchical relationships among the clusters found. Of the remaining four phenograms, two were intermediate between patterns A and C while the other two could not be categorized. Formal multivariate methods of phenogram comparison are available (Podani, 1982; Section 6.3), but these were not implemented here since they are computationally impractical with large samples. Instead, the hierarchical relationships of groups of OTUs in the phenograms produced here have been compared in the preceding section in groups of four, one for each of the three sorting algorithms used. In the present section further comparisons will be made of all twelve phenograms. These will concentrate on the behavior of particular groups of OTUs and on the apparent effect of particular descriptors and of the resemblance coefficients employed on the results obtained.

5.4.1 Group Structure

In addition to the recovery of the topodeme samples as clusters, five other groups of OTUs behaved fairly consistently throughout the analyses reported here. These were the following: (a) five 20-stamen crus-galli OTUs (204, 713, 714, 751, 752); (b) four 10-stamen crus-galli OTUs with red anthers not from Site 4 (702, 791-793); (c) the sample of seven OTUs of C. ?grandis; (d) the sample of eleven OTUs of C. sp. aff. C. bushii; and (e) the three OTUs belonging to C. section Macracanthae (761, 762, 769).

The first of these is a group of large-fruited OTUs. At Sites 2 and 11, where the comparison has been made, OTUs 751 and 752 are distinguished by flowering noticeably earlier than other, sympatric 20-stamen C. crus-galli. These five OTUs, accompanied frequently by OTUs 753 and 754, occur together in all but four phenograms. In Figures 5.2 and 5.6 OTU 752 is detached from the group. In figure 5.3 and 5.4, all five OTUs are dispersed in the region between the C. crus-galli and C. punctata OTUs.

The second group of OTUs is distinguished not only by its red anthers but also by the presence of the projection of the stamen connective. This is the only instance of this among 10-stamen C. crus-galli OTUs encountered in this study. The resulting intermediacy of these OTUs with

respect to the 10- and 20-stamen morphotypes of C. crus-galli appears to account for their association with the 20-stamen OTUs in some instances (branch D in Figures 5.5 and 5.9).

The samples of both C. ?grandis and C. sp. aff. C. bushii are so distinct in all the analyses that they are rarely subdivided or mingled with other OTUs in any way. The exceptions that occur are when OTUs of C. punctata and C. ?grandis form a common cluster (Fig. 5.7, K; Fig. 5.11, J) or when the OTUs of C. macracantha and C. succulenta (C. section Macracanthae) are joined to those of C. bushii (Fig. 5.11, H; Fig. 5.12, F). In the first case, the association of these two taxa is due to the fact that features distinguishing C. ?grandis and C. punctata (e.g. leaf glossiness) are not included among the descriptors used here. The second case is due to the fact that the bushii and Macracanthae OTUs are the only ones in the sample showing such a consistent and extreme degree of calyx lobe toothing. This affinity with the bushii OTUs notwithstanding, however, in most cases the Macrancathae ones are those most unrelated to the rest of the sample, often forming a distinct out-group (Fig. 5.5, L; Fig. 5.6, M; Fig. 5.8, L; Fig. 5.10, M).

5.4.2 Effect of Individual Descriptors

From the preceding sections it is evident that particular descriptors appear to have been especially important in determining the group structures seen in Figures 5.1 through 5.12. The mechanism of this importance probably varies. The two descriptor ranking methods employed in Section 3.7.3 distinguish descriptors which are highly correlated with other descriptors (Table 3.7a) from those which are less correlated but convey unique information (Table 3.7b). It appears that both of these features are involved in the way in which the phenograms obtained here reflect the distribution of particular descriptor states.

All twelve phenograms demonstrate a sorting of the sample according to stamen number (STAM) and style number (STYL). However, except for the correlations STYL-TCAL, STYL-ANTH, STAM-TCAL and STAM-LCAL the correlations between style number and stamen number, and between these two and the other descriptors are largely explainable by correlations among the other descriptors, judging from the results of descriptor ranking (Table 3.7) and comparisons of correlation and partial correlation coefficients (Table 3.8; Rohlf, 1977). Thus contrasts between groups of OTUs with respect to their scores for STYL and STAM are reinforced by their scores on most other descriptors.

On the other hand, it is also evident that descriptors with quite low redundancy such as LCAL, ANTH and TCAL (Table 3.7b) have also played a major part in distinguishing particular groups of OTUs. The distinctness of the sample of C. ?grandis, and that of topodeme sample T6 appear to be due in part to high and low scores, respectively, for LCAL. The association of high scores on the two multistate descriptors TCAL and ANTH with conspicuously distinct groups of OTUs has already been referred to above (Section 5.4.1). Their scores for ANTH contribute to distinguishing T6 and the other group of red-anthered OTUs (702, 791-793). The OTUs of C. sp. aff. C. bushii and C. section Macracanthae are distinct by virtue of their high scores on TCAL.

5.4.3 Comparison of Resemblance Coefficients

The twelve phenograms reported here were obtained from four matrices of resemblance coefficients. The euclidean distance matrices (E, *E) and the matrix of information radii (D) all gave similar results with single- and average-linkage clustering (patterns A or C, or their intermediates). The matrix of generalized distances (M) gave quite different results with these two algorithms, consonant with the relatively low correlation between M and the other three matrices (Table 4.1). Only with the

extremely space-dilating minimum variance method did M produce a pattern of hierarchical relationshps in common with any of the other resemblance coefficients (pattern A). The significance of these and similar results obtained with the 60 OTU subsample are discussed further in Section 6.4.2.

5.4.4 Interaction of Descriptor and Resemblance Coefficient

The contribution of multistate descriptors to the analyses reported here was especially pronounced in the case of those based on the euclidean distances calculated from ranged data (E). This is probably due in large part to the use of global maxima and minima in ranging the data (Section 3.6, Table 3.1). Because the multistate descriptors were often invariant within OTUs as well as within topodeme samples, their mean scores for an OTU could often correspond to global maxima or minima (e.g. ANTH and OTUs with red anthers versus those with ivory ones; PROJ and the OTUs of 20-stamen C. crus-galli; PUB1 and OTUs of C. punctata versus those of T1, T3, T4 and T6; Table 3.11). In constrast, none of the OTU means of the continuous and meristic descriptors correspond to global maxima or minima, since these descriptors were almost never invariant. Consequently the values of these descriptors in the n x p matrix R were in a much narrower range than those of some of the multistate ones.

Although it was done to ensure both commensurability and equal weighting of the flower and fruit descriptors, ranging by global maxima and minima had the effect of systematically weighting certain multistate descriptors over all the rest. The consequences of this may be seen in the very pronounced group structure of the three phenograms based on $\underset{\sim}{E}$ (Fig. 5.1, 5.5, 5.9), as discussed above. However, re-calculating $\underset{\sim}{E}$ from a data matrix in which all eleven descriptors varied in the 0 - 1.0 range (as described for the leaf data in Section 3.6), and repeating the three cluster analyses (not shown) gave results that differed little from those presented already. Both single- and average-linkage clustering produced pattern C, while minimum variance clustering produced pattern A.

5.4.5 Conclusions

The results presented in this chapter have demonstrated a number of features of the flower and fruit data obtained for the entire 160 OTU sample. The data unquestionably support subdivision of the sample into a number of distinct groups of similar OTUs which correspond in part to the individual topodeme samples in the total sample and also to the taxa recognized at the outset (Table 5.1). This in itself is of interest since differentiation in C. section Crus-galli has not been demonstrated in this way previously.

either at the local level (topodemes) or at the level of taxa of whatever rank (morphotypes).

At the same time, some less obvious aspects of the data were also indicated. One is the apparent degree of differentiation of the 10-stamen C. crus-galli with red anthers from other C. crus-galli. Another is the intermediacy of the five 20-stamen C. crus-galli OTUs (204, 713, 714, 751, 752), with respect to C. crus-galli sensu lato and C. punctata (Fig. 5.2-5.4, 5.6, 5.7).

However, the results obtained have also suggested that much of the group structure observed in Figures 5.1 - 5.12 may be due largely to particular descriptors or data types. The investigations described in the following three chapters are concerned with further investigating this possibility, and with attempting to evaluate the indications of group structure in the sample that have been obtained here.

CHAPTER SIX

DETERMINANTS OF GROUP STRUCTURE

6.1 Introduction

The preceding chapter has reported the results of cluster analyses of the total 160 OTU sample. These analyses were based on resemblance matrices calculated (Chapter 4) from the flower and fruit data (Table 3.1) for all OTUs. The present chapter is concerned with an exploration of how different descriptors contribute to the group structure found in the cluster analyses.

In addition to the flower and fruit data, leaf data were available for a subsample of 60 OTUs (Table 3.2). The contribution of individual descriptors to the classification of OTUs is examined both through component analyses of the two data sets (160 OTU and 60 OTU; Section 6.2) and by the comparison of additional classifications of the 60 OTU subsample (Section 6.3). The latter are based on the resemblance matrices calculated in Chapter 4 from different combinations of flower, fruit and leaf descriptors (Table 4.2).

6.2 Component Analyses

Methods of component analysis are based on eigenanalysis of n x n or p x p matrices. In the case of the p x p matrix **RR** of descriptor correlations (or the corresponding variance-covariance matrix **S**), the R algorithm of principal components analysis (R-PCA) is employed. Alternatively, it may be computationally simpler to use the n x n matrix **Q** of scalar products between OTU vectors as the basis for eigenanalysis, in which case the Q algorithm is employed (as in Section 4.5). Similarly, the OTUs may also be ordinated following eigenanalysis of the n x n matrix **DD** of resemblances between OTUs; this is principal coordinates analysis (PCoA; Gower, 1966). Each of these three methods results in partition of the total variation in the sample among successive orthogonal axes. This allows a summary of the data structure to be visualized in a space defined by the first few of these axes.

Component analyses were carried out using the programs given by Orlóci (1978; PCAR, PCAQ, PCAD) as well as program PRINCOOR written by J. Podani. Correlations between individual principal components and the original descriptors were calculated for R-PCAs according to the formulae in Pimentel (1979). The percent of each descriptor's variance accounted for by each component, plus Bartlett's approximate test for the equality of sets of eigenvalues were also

calculated (Pimentel, 1979). Results of component analyses are displayed either as graphs of selected pairs of components, or by means of stereo diagrams representing three components at once (Programs PREP, STEREO, and HPSTPL; Fewster & Orlóci, 1978).

PCA methods make a number of assumptions about the origin and structure of the data analyzed. First, it is assumed that the data approximate a linear, multivariate normal distribution with finite centroid and finite covariance matrix (Pimentel, 1979). Second, but related to the first, is the assumption that the data is continuous and that the sampling units are drawn from a common population. With regard to the six descriptors analyzed in detail in Section 3.7 (STYL, TCAL, WFL, LCAL, LFR, WFR) these assumptions appeared likely to be justified, for the most part, as their univariate distributions were normal or nearly so, and their bivariate distributions were largely linear and continuous. However, the distribution of stamen number (STAM) in the sample was bimodal. Distributions of the remaining binary (PROJ) and multistate descriptors (ANTH, PUB1, PUB2) were markedly discontinuous and non-normal (Table 3.11). Although Pimentel (1979) suggests that the centroids of large samples tend to approach multivariate normality, it seems as if this might be unreasonably optimistic in the case of all eleven flower and

fruit descriptors taken together.

Nevertheless, it was considered informative to carry out component analyses on the data in order to observe the consequences of departures from multivariate normality. Similarly, if the sample examined proved not to be homogeneous, this would validate the results of cluster analyses done on the same data. In addition, it can be shown that the R-PCA plot of a sample calculated from its correlation matrix $\underset{\sim}{R}$ is equivalent to that produced by PCoA of a matrix of euclidean distances $^*\underset{\sim}{E}$ calculated from the same data (Gower, 1966). In this way the behavior of OTUs in cluster analyses based on $^*\underset{\sim}{E}$ can be related, via the R-PCA, to their scores on individual (standardized) descriptors.

6.2.1 160 OTU Sample

Two R-PCAs of the 160 OTU sample were done based on the eleven flower and fruit descriptors used in calculating the resemblance matrices ($\underset{\sim}{E}$, $^*\underset{\sim}{E}$, $\underset{\sim}{M}$, $\underset{\sim}{D}$) which were the basis of the cluster analyses reported in Chapter 5. The first of these used the covariance matrix of the ranged descriptors (Fig. 6.1a), while the second used the corresponding correlation matrix (Fig. 6.2a). Both analyses successfully reduced the dimensionality of the data, the first three components accounting for 85.2% and 74.7% of the total

variation in the covariance and correlation analyses respectively (Table 6.1, 6.2).

The ordinations produced by these two analyses differ markedly (compare Fig. 6.1a and 6.2a). In Figure 6.1a the OTUs of C. punctata and C. ?grandis form a loose cluster of points at the right of the plot (3, in Fig. 6.1a), while the crus-galli ones are separated principally into two clusters, at the lower left (1, in Fig. 6.1a; 10-stamen C. crus-galli plus C. sp. aff. C. bushii) and one at the upper left (2, in Fig. 6.1a; 20-stamen C. crus-galli). The individual topodeme samples of 10- and 20-stamen C. crus-galli (T1 - T6) form very dense clusters of OTUs. In Figure 6.2a the clusters representing these samples are more diffuse, but remain distinct. As in Figure 6.1a the OTUs of C. punctata and C. ?grandis (3, in Fig. 6.2a) are at one side of the plot, while those of 10-stamen C. crus-galli (1, in Fig. 6.2a) are at the opposite side. The OTUs of both 20-stamen C. crus-galli (2, in Fig. 6.2a) and C. sp. aff. C. bushii (4, in Fig. 6.2a) lie between the clusters of punctata and 10-stamen crus-galli OTUs.

Examination of the component-descriptor correlations and eigenvector elements for these two analyses showed that the first components of the covariance matrix were dominated by the descriptors PROJ, PUB1, PUB2 and to a lesser extent, ANTH (Fig. 6.1b; Table 6.1). This is in contrast to the

correlation-based analysis in which no one descriptor stood out as contributing appreciably more to the first two or three components than any other (Fig. 6.2b; Table 6.2).

Since the descriptors in the covariance-based PCA had been ranged so as to make them commensurate, the disparity among the descriptors with respect to their influence on the analysis was unexpected. However, the following observations were made. The descriptors contributing most to the first two components are all binary or multistate ones shown earlier to be discontinuously distributed (Section 3.7; Table 3.11) as well as over-weighted by ranging as it was carried out using global maxima and minima (Section 3.6, 5.4.4). As a consequence these descriptors have exaggerated variances and covariances (Fig. 6.3; Table 6.1). Since these descriptors also tend to be invariant within topodeme samples and morphotypes they are probably largely responsible for the very tight clustering of OTUs seen in Figure 6.1a.

Repeating the analysis using a covariance matrix calculated after ranging all eleven flower and fruit descriptors to a uniform 0 - 1.0 range (as described for the leaf data in Section 3.6; not shown) gave quite similar results to those described in Figure 6.1 and Table 6.1, however. The individual topodeme samples of the two common morphotypes of C. crus-galli were somewhat more spread out

in the plane of the first two principal components, but otherwise the essential features of Figure 6.1a remained. The second ranging increased the variances of the continuous and meristic descriptors in Table 6.1 two- or three-fold, while leaving those of the binary and multistate ones unchanged. Similarly, some of the covariances (PROJ-STAM, PUB1-PUB2) also remained appreciably larger than the rest. Another feature in common to these two covariance-based PCAs and distinguishing them from the correlation matrix analysis was the greater dimensionality of the sample. Although the first three covariance components account for 85.2% of the total variation, the 7.4% accounted for by the fourth component represents 86.2% of the variance of TCAL, the descriptor largely responsible for the separation of the OTUs of C. sp. aff. C. bushii, C. macracantha and C. succulenta from the remainder of the sample (Section 5.4.2; compare also Fig. 6.1 and 6.2). and 6.2a and b).

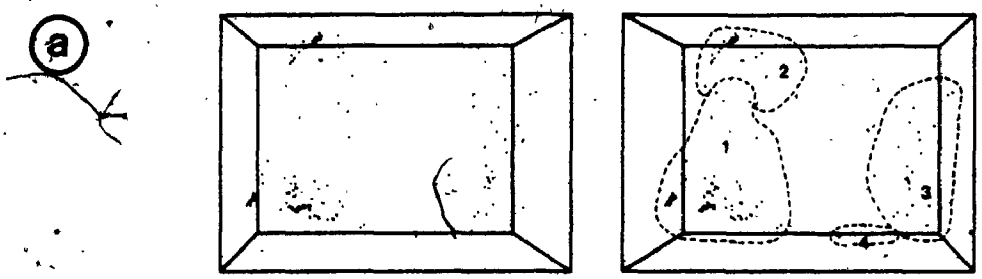
Obviously, the correlation-based PCA provides a more efficient summarization of the flower and fruit data than does the covariance one since more of each descriptor's variance is accounted for by the first three components in this analysis. At the same time, however, it must be recognized that the standardization of the scores for each descriptor to zero mean and unit standard deviation implied by R-PCA of the correlation matrix imposes a new dispersion

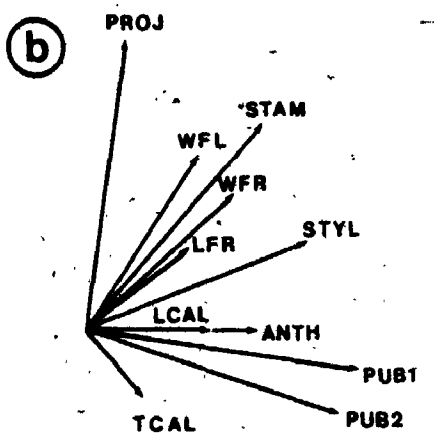
structure on the data. Orlóci (1978) has pointed out that the process of ranging leaves intact the original variance-covariance structure of the data. In using PCA methods for data exploration this feature of ranging appears to be a very useful one, particularly in conjunction with the opportunity to compare covariance- and correlation-based results.

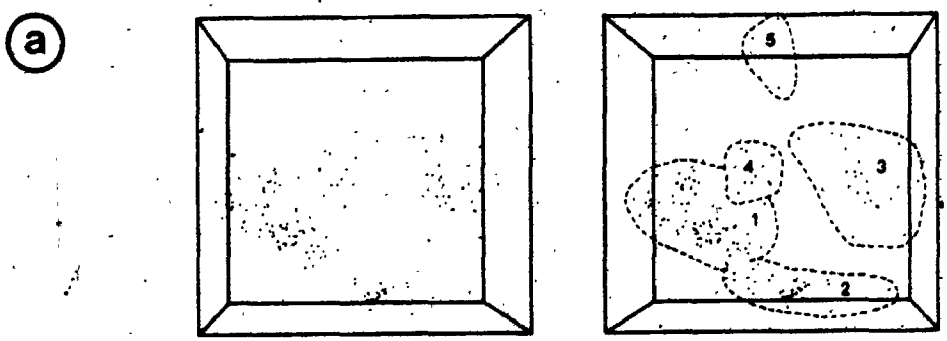
The differences observed here between the two dispersion structures of the 160 OTU sample, and between the binary and multistate descriptors on the one hand, and the continuous and meristic ones on the other correspond to the differences observed among the cluster analyses reported in Chapter 5. In particular, OTU scores for PROJ, PUB1, PUB2, and ANTH were probably responsible for the very low levels at which the members of the same crus-galli topodeme sample fused in the analyses based on $\underset{\sim}{E}$ (euclidean distances calculated from ranged data; Fig. 5.1, 5.5, 5.9), as compared with those based on $*\underset{\sim}{E}$ (euclidean distances calculated from standardized data; Fig. 5.2, 5.6, 5.10). If taxonomic decisions are to be based to any extent on numerical analyses such as those described here and in Chapter 5, it is extremely important to be aware of the effects that particular combinations of data type and method may have on the results. These considerations are especially relevant also to the comparisons made in the following sections and in Chapter 7 and 8 as well.

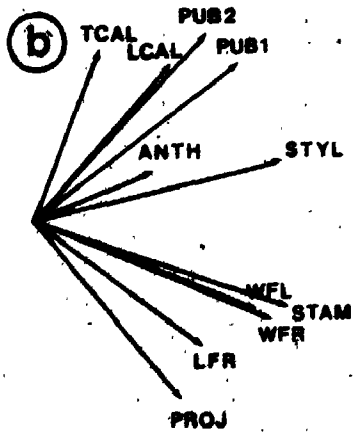
Figure 6.1 R-algorithm principal components analysis of the 160 OTU sample, based on the covariance matrix for eleven ranged flower and fruit descriptors. In (a), stereo diagram of the sample in the space defined by the first three principal components (1, 10-stamen C. crus-galli plus C. sp. aff. C. bushii; 2, 20-stamen C. crus-galli; 3, C. punctata plus C. ?grandis; 4, C. Section Macracanthae); (b) vector diagram of the correlations between individual descriptors and the first two component axes. See text and Table 6.1 for details.

Figure 6.2 R-algorithm principal components analysis of the 160 OTU sample, based on the correlation matrix for eleven flower and fruit descriptors. In (a), stereo diagram of the sample in the space defined by the first three principal components (1, 10-stamen C. crus-galli; 2, 20-stamen C. crus-galli; 3, C. punctata plus C. ?grandis; 4, C. sp. aff. C. bushii; 5, C. Section Macracanthae); (b) vector diagram of the correlations between the individual descriptors and the first two component axes. See text and Table 6.2 for details.


a
1
2
3
4


b
PROJ
STAM
WFL
WFR
LFR
STYL
LCAL
ANTH
PUB1
PUB2
TCAL


a
1
2
3
4
5


b
TCAL
LCAL
PUB2
PUB1
ANTH
STYL
WFL
STAM
WFR
LFR
PROJ

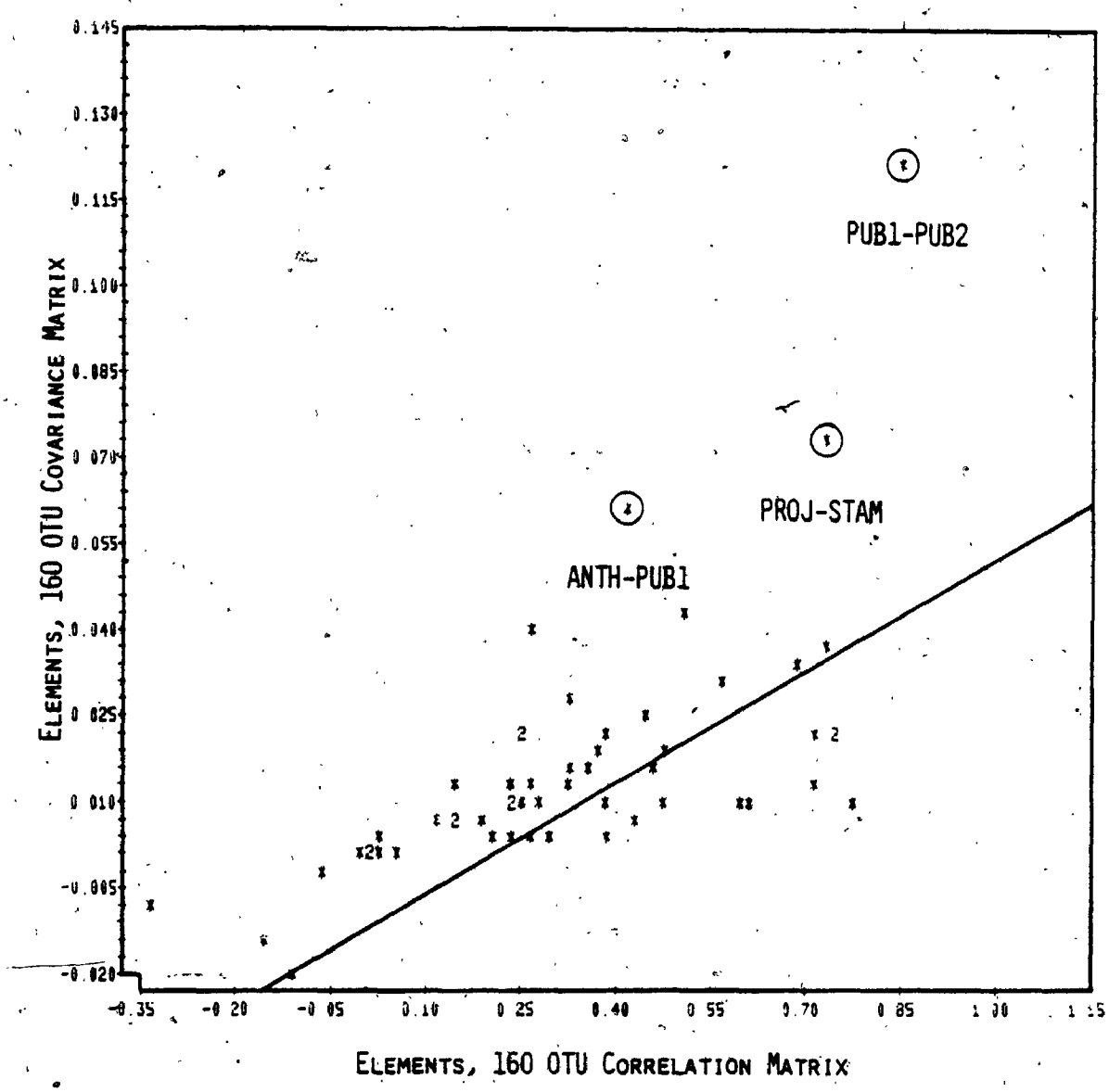


Figure 6.3 160 OTU sample, regression of the elements of the covariance matrix for 11 ranged flower and fruit descriptors on the elements of the corresponding correlation matrix. Circled points, representing the elements indicated, have larger residuals than expected (standardized residuals > 2.0; Ryan, Joiner & Ryan, 1976). See Table 3.1 for explanation of descriptor abbreviations.

Table 6.1 Parameters of the first four components of the R-PCA of the 160 OTU sample based on the variance-covariance matrix for 11 ranged flower and fruit descriptors. Eigenvalues underlined are significantly different from those to the right ($p < 0.001$, except for IX where $0.01 < p < 0.05$) according to Bartlett's sphericity test (Pimentel, 1979).

	I	II	III	IV	V	VI	VII	VIII	IX
Eigenvalue	0.334	0.228	0.125	0.060	0.023	0.018	0.007	0.005	0.003
% Trace	41.4	28.3	15.5	7.4	2.8	2.2	0.9	0.6	0.4
Cumulative %		69.7	85.2	92.6	95.3	97.6	98.5	99.1	99.5

		I	II	III	IV	I	II	III	IV
Des(1)	Variance(2)	Correlations(3)				Eigenvector Elements(4)			
STYL	0.017	0.773	0.275	-0.400	0.122	0.175	0.076	-0.149	0.066
STAM	0.053	0.628	0.686	-0.088	0.010	0.249	0.329	-0.057	0.010

Table 6.1 Cont.

		I	II	III	IV	I	II	III	IV
Des(1)	Variance(2)	Correlations(3)				Eigenvector Elements(4)			
PROJ	0.182	0.164	0.973	-0.056	0.039	0.121	0.869	-0.067	0.069
TCAL	0.061	0.197	-0.230	-0.196	0.928	0.084	-0.119	-0.137	0.936
PUB1	0.142	0.937	-0.150	-0.158	-0.118	0.611	-0.118	-0.169	-0.183
WEL	0.017	0.412	0.580	0.030	0.293	0.093	0.159	0.011	0.157
ECAL	0.009	0.529	-0.080	-0.015	0.306	0.088	-0.016	-0.004	0.121
LFR	0.011	0.363	0.269	0.264	-0.341	0.066	0.060	0.079	-0.147
WFR	0.017	0.526	0.452	0.006	-0.087	0.119	0.123	0.002	-0.047
PUB2	0.143	0.860	-0.299	-0.327	-0.049	0.564	-0.237	-0.351	-0.075
ANTH	0.154	0.588	-0.052	0.802	0.067	0.399	-0.042	0.891	0.108

(1) Descriptor (see Table 3.1 for explanation of descriptor abbreviations).

(2) Descriptor variance for the 160 OTU sample.

(3) Descriptor-component correlations.

(4) Columns I-IV represent the eigenvectors corresponding to the first four principal components of the 160 OTU variance-covariance matrix.

Table 6.2 Parameters of the first five components of the R-PCA of the 160 OTU sample based on the correlation matrix for 11 ranged flower and fruit descriptors. Eigenvalues underlined are significantly different from those to the right ($p < 0.001$) according to Bartlett's sphericity test (Pimentel, 1979).

	I	II	III	IV	V	VI	VII	VIII	IX
Eigenvalue	<u>4.684</u>	<u>2.184</u>	<u>1.351</u>	<u>0.961</u>	<u>0.662</u>	<u>0.524</u>	<u>0.219</u>	0.151	0.110
% Trace	42.6	19.9	12.3	8.7	6.0	4.8	2.0	1.4	1.0
Cumulative %		62.4	74.7	83.4	89.5	94.2	96.2	97.6	98.6

	I	II	III	IV	V	I	II	III	IV	V
Des(1)	Correlations(2)					Eigenvector Elements(3)				
STYL	0.863	0.190	0.237	-0.305	0.044	0.399	0.128	0.204	-0.311	0.055
STAM	0.871	-0.299	0.177	-0.112	-0.189	0.402	-0.203	0.152	-0.114	-0.232
PROJ	0.503	-0.600	0.408	-0.034	-0.388	0.233	-0.406	0.351	-0.035	-0.477

Table 6.2 Cont.

	I	II	III	IV	V	I	II	III	IV	V
Des(1)	Correlations(2)					Eigenvector Elements(3)				
TCAL	0.140	0.577	0.572	0.353	0.271	0.065	0.390	0.492	0.360	0.333
PUB1	0.722	0.523	-0.232	-0.262	-0.162	0.333	0.354	-0.199	-0.267	-0.199
WFL	0.759	-0.315	0.306	0.285	0.138	0.351	-0.213	0.264	0.291	0.170
LCAL	0.488	0.516	0.099	0.384	-0.066	0.225	0.349	0.085	0.392	-0.082
LFR	0.579	-0.427	-0.547	0.087	0.333	0.268	-0.289	-0.471	0.089	0.410
WFR	0.826	-0.340	-0.140	0.035	0.327	0.382	-0.230	-0.120	0.036	0.402
PUB2	0.614	0.635	-0.175	-0.332	0.031	0.284	0.429	-0.150	-0.338	0.038
ANTH	0.424	0.158	-0.513	0.560	-0.363	0.196	0.107	-0.441	0.571	-0.446

(1) Descriptor (see Table 3.1 for explanation of descriptor abbreviations).

(2) Descriptor-component correlations.

(3) Columns I-V represent the eigenvectors corresponding to the first five principal components of the 160 OTU correlation matrix.

6.2.2 60 OTU Subsample

As described in Chapter 3, data for a total of 17 descriptors were available for a subsample of 60 of the randomly sampled OTUs (Section 3.5, Table 3.2). These data were used to investigate the contributions made by different subsets of descriptors to the group structure of the subsample, as revealed by classification and component analyses. The classification analyses are described in Section 6.3. The remainder of this section is concerned with the ordination results.

Four contrasting data sets were prepared from the 60 x 17 data matrix by forming the following subsets of descriptors as described in Section 4.5.2 (Table 4.2): flower and fruit (PROJ, ANTH, TCAL, PUB1, PUB2, WFL, LCAL, LFR, WFR) versus leaf (X, Y, Z, NUMSEC, ANGSEC, TEETH) and binary and multistate (PROJ, ANTH, TCAL, PUB1, PUB2, NUMSEC, ANGSEC) versus continuous and meristic (WFL, LCAL, LFR, WFR, X, Y, Z, TEETH). Principal component analyses based on each of these subsets were compared with a PCA using all 17 descriptors and with PCoA of the matrix of the information radii (INF) calculated for the 60 OTU subsample from all eleven flower and fruit descriptors.

The resulting six ordinations of the 60 OTU subsample are displayed in Figures 6.4 through 6.9. Tables 6.3

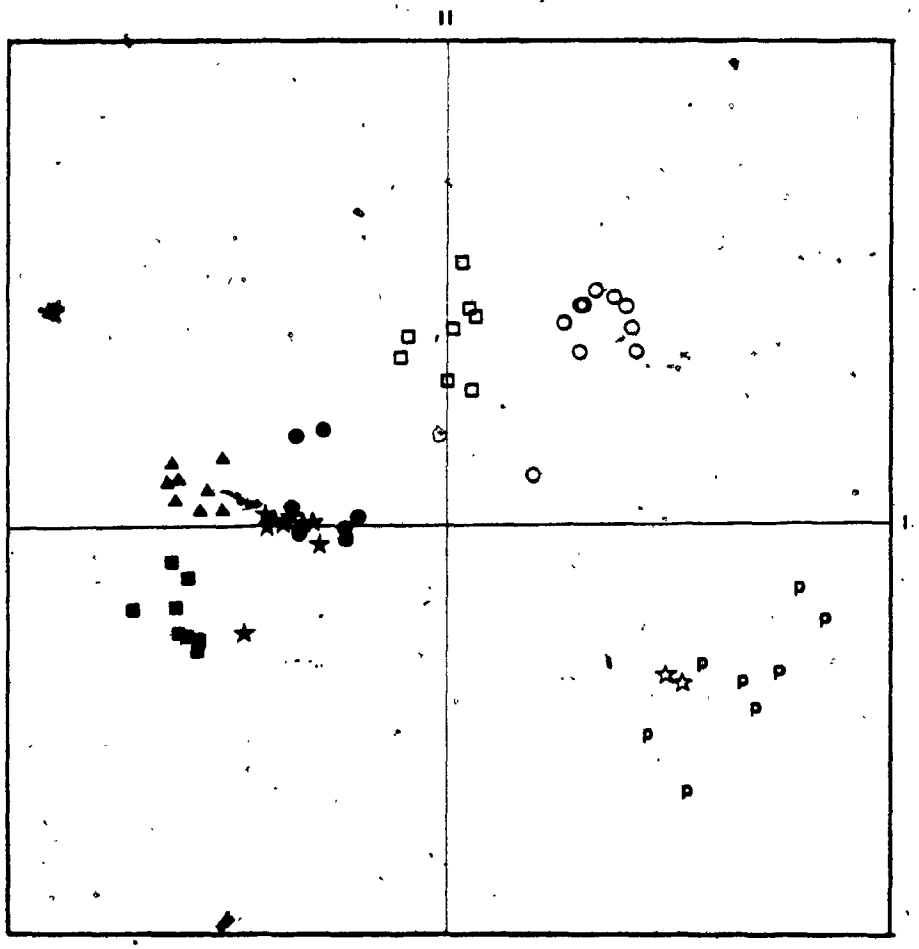
through 6.8 provide details of these analyses. For the first three components of each of the five PCAs (accounting for 77.8 - 90.8% of the total variation; Table 6.3-6.8), both the component-descriptor correlations and the eigenvector elements are given. The former provide a probabilistic description of descriptor contributions to each analysis, while the latter describe deterministically how the OTU configurations seen in the figures were produced.

The following points are evident from examination of the six ordinations of the 60 OTU subsample. The differentiation of the topodeme samples is strongest in the ordinations based on some or all of the flower and fruit descriptors, and is a function more of the multistate descriptors than of the continuous ones (Fig. 6.5; also compare Fig. 6.7 and 6.8). The latter point lends support to inferences made in Chapter 5 concerning the role of these descriptors in the cluster analyses. Note that here, unlike in the analyses of the 160 OTU sample described in Chapter 5 and in the preceding section, the descriptors have been ranged uniformly, using the maximum and minimum OTU means for this purpose rather than the global maxima and minima (Section 3.6, 4.5.2). Consequently the effects of the continuous and meristic descriptors on the analyses are not artificially reduced in comparison.

Differentiation of topodeme samples is present in the ordinations based on the continuous descriptors only (Fig. 6.8) and on the leaf descriptors only (Fig. 6.9), but in these, especially the latter, it is quite weak. Note that the differentiation exhibited in the R-PCA of the leaf data is principally between punctata and crus-galli OTUs with high stamen numbers on one hand, and 10-stamen crus-galli on the other. (Fig. 6.10).

On grounds of the heterogeneity of the sample it may be objected that PCA is an unsuitable method for the comparison of the effects different descriptors have in determining group structure. The canonical analyses described in Chapter 8 are better suited, and provide a quantitative description of the degree to which a priori groups of OTUs are distinguishable by means of the descriptors used. However, the purpose in making these component analyses was to summarize the data without making any assumptions about their structure. The implications of the results obtained here with respect to the contribution of particular descriptors will be seen in the following section where classifications based on the same data sets will be compared.

Figure 6.4 60 OTU subsample in the space defined by the first two principal coordinates of the matrix of information radii (INF) among the OTUs. INF is calculated on the basis of all 11 flower and fruit descriptors. The 10-stamen morphotype of C. crus-galli is represented by solid symbols, the 20-stamen one by open symbols, and C. punctata (Site 5; topodeme sample T7) by the letter p. Individual sites are coded by geometric shape: squares (Site 1; topodeme samples T1, T2); circles (Site 2; topodeme samples T4, T5); triangles (Site 3; topodeme sample T3); and stars (Site 4; topodeme sample T6). Open stars indicate OTUs of C. ?grandis (Site 4).


II
I

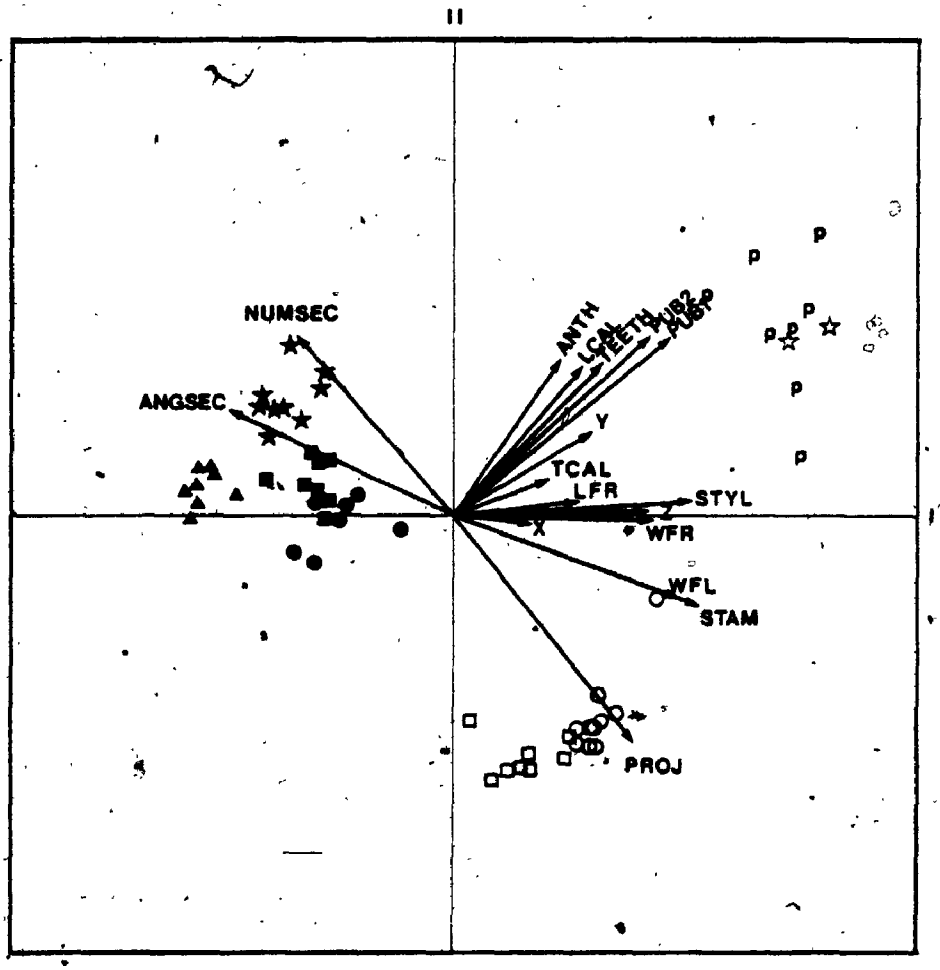


Figure 6.5 60 OTU subsample in the space defined by the first two principal components of the variance-covariance matrix based on 17 ranged flower, fruit and leaf descriptors. Vectors represent the relative magnitudes of component-descriptor correlations in this space. See text and Table 6.3 for details. See Table 3.1, 3.3 for explanation of descriptor abbreviations. Symbols as in Figure 6.4.

Table 6.3 Parameters of the first three components of the R-PCA of the 60 OTU subsample based on the covariance matrix for 17 ranged flower, fruit and leaf descriptors (see Table 3.1, 3.3 for explanation of descriptor abbreviations).

	Components							
	I	II	III	IV	V	VI	VII	VIII
Eigenvalue	0.56	0.32	0.18	0.09	0.05	0.04	0.03	0.02
% Trace	41.5	23.5	13.3	6.8	3.7	2.8	2.1	1.3
Cumulative %		65.0	78.3	85.1	88.8	91.6	93.7	95.0

	I	II	III	I	II	III
Descriptor	Correlations			Eigenvector Elements		
STYL	0.898	0.051	-0.284	0.315	0.024	-0.176
STAM	0.938	-0.305	0.052	0.473	-0.204	0.046
PROJ	0.595	-0.760	0.026	0.344	-0.584	0.027
TCAL	0.319	0.118	-0.702	0.090	0.044	-0.349
PUB1	0.733	0.606	-0.221	0.351	0.385	-0.187
WFL	0.762	-0.285	0.315	0.242	-0.120	0.176
LCAL	0.433	0.494	-0.156	0.123	0.186	-0.078
LFR	0.421	0.049	0.637	0.126	0.019	0.337
WFR	0.675	-0.011	0.345	0.215	-0.005	0.194
PUB2	0.657	0.605	-0.297	0.312	0.382	-0.249
ANTH	0.356	0.527	0.633	0.192	0.378	0.602

Table 6.3 Cont.

	I	II	III	I	II	III
Descriptor	Correlations			Eigenvector Elements		
X	0.262	-0.035	0.507	0.082	-0.015	0.279
Y	0.470	0.281	0.482	0.111	0.088	0.200
Z	0.654	0.011	0.117	0.152	0.004	0.048
NUMSEC	-0.523	0.594	0.098	-0.151	0.228	0.050
ANGSEC	-0.766	0.352	0.307	-0.263	0.161	0.186
TEETH	0.498	0.515	-0.364	0.151	0.208	-0.195

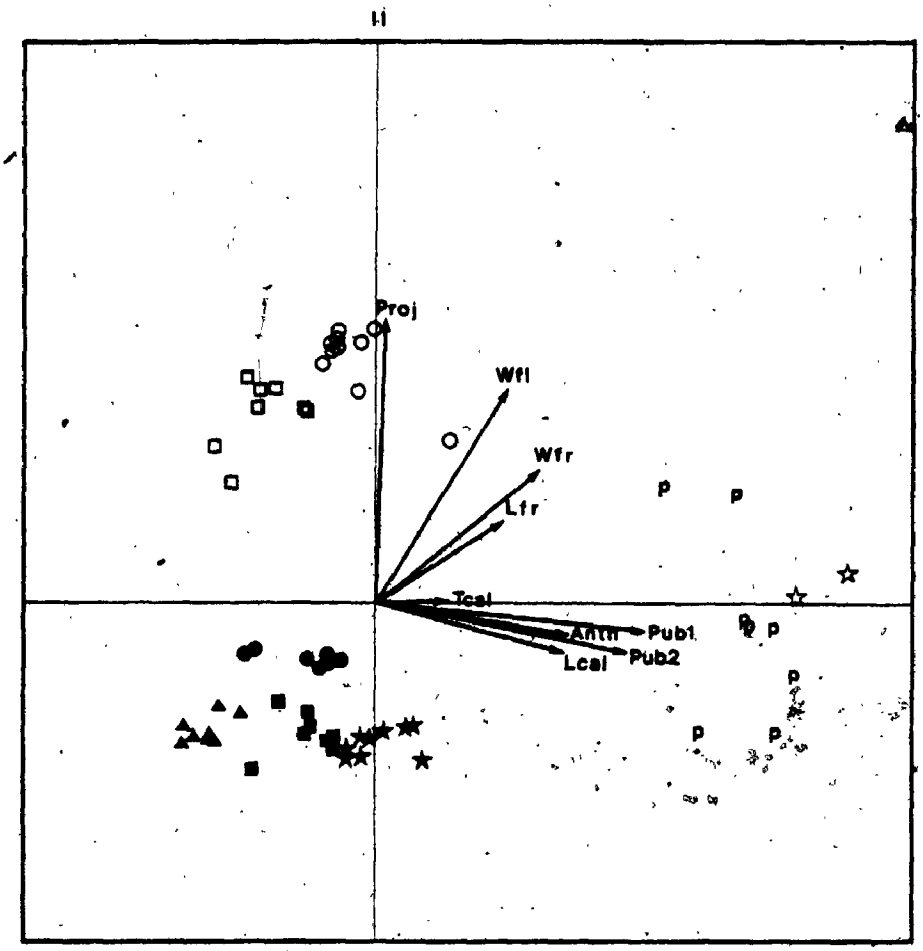


Figure 6.6 60 OTU subsample in the space defined by the first two principal components of the variance-covariance matrix based on 9 ranged flower and fruit descriptors. Vectors represent the relative magnitudes of component-descriptor correlations in this space. See text and Table 6.4 for details. See Table 3.1 for explanation of descriptor abbreviations. Symbols as in Figure 6.4.

Table 6.4 Parameters of the first three components of the R-PCA of the 60 OTU subsample based on the covariance matrix for 9 ranged flower and fruit descriptors (see Table 3.1 for explanation of descriptor abbreviations).

	Components							
	I	II	III	IV	V	VI	VII	VIII
Eigenvalue	0.34	0.23	0.15	0.07	0.03	0.02	0.01	0.01
% Trace	39.1	26.9	16.9	7.9	3.2	2.3	1.6	1.2
Cumulative %		66.0	82.9	90.8	94.0	96.3	97.9	99.1

	I	II	III	I	II	III
Descriptor	Correlations			Eigenvector Elements		
PROJ	0.030	0.973	-0.104	0.022	0.876	-0.118
TCAL	0.253	0.007	-0.705	0.092	0.003	-0.391
PUB1	0.927	-0.090	-0.287	0.573	-0.067	-0.270
WFL	0.452	0.732	0.151	0.185	0.361	0.094
LCAL	0.643	-0.161	-0.166	0.236	-0.071	-0.093
LFR	0.431	0.287	0.472	0.167	0.134	0.278
WFR	0.564	0.459	0.174	0.232	0.228	0.109
PUB2	0.860	-0.167	-0.412	0.528	-0.124	-0.384
ANTH	0.663	-0.107	0.671	0.462	-0.089	0.711

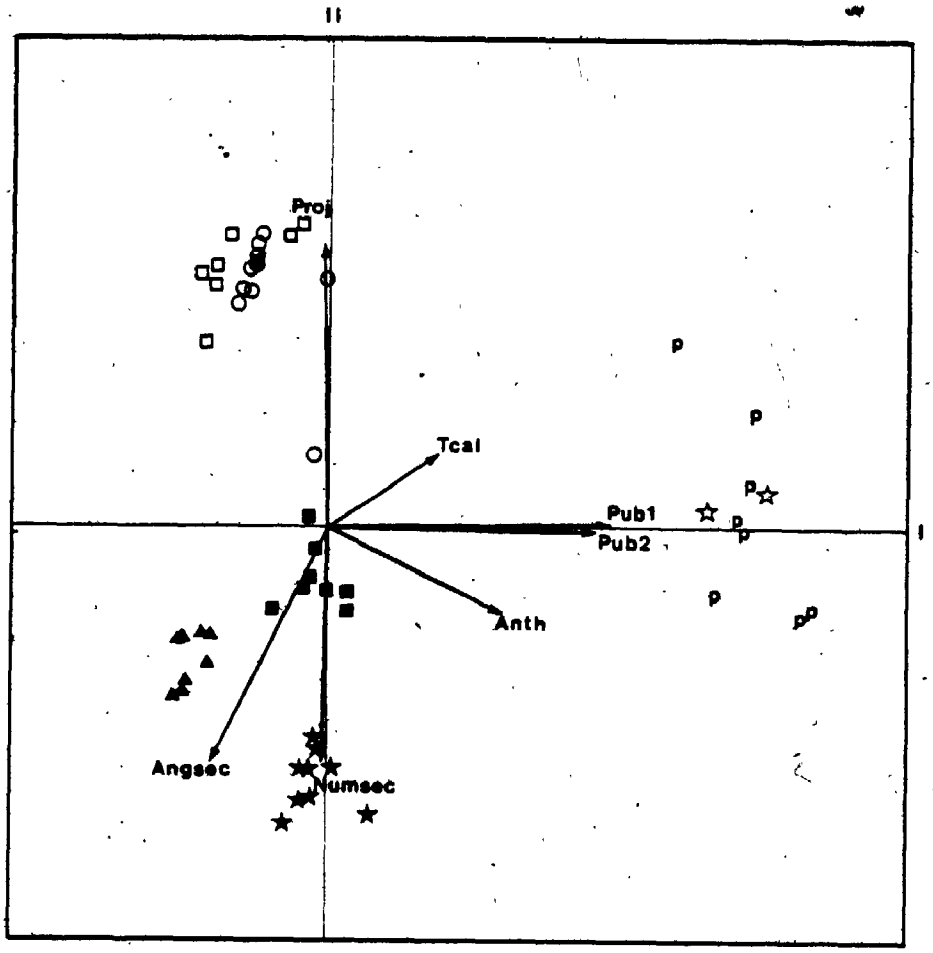


Figure 6.7 60 OTU subsample in the space defined by the first two principal components of the variance-covariance matrix based on 7 ranged multistate flower, fruit and leaf descriptors. Vectors represent the relative magnitudes of component-descriptor correlations in this space. See text and Table 6.5 for details. See Table 3.1 and 3.3 for explanation of descriptor abbreviations. Symbols as in Figure 6.4.

Table 6.5 Parameters of the first three components of the R-PCA of the 60 OTU subsample based on the covariance matrix for 7 ranged multistate flower, fruit and leaf descriptors (see Table 3.1, 3.3 for explanation of descriptor abbreviations).

	Components						
	I	II	III	IV	V	VI	VII
Eigenvalue	0.30	0.26	0.13	0.03	0.02	0.01	0.01
% Trace	39.3	34.2	17.3	3.7	2.6	1.6	1.3
Cumulative %		73.5	90.8	94.5	97.1	98.7	100.0

	I	II	III	I	II	III
Descriptor	Correlations			Eigenvector Elements		
PROJ	-0.014	0.956	0.246	-0.011	0.810	0.293
TCAL	0.378	0.253	-0.526	0.146	0.105	-0.306
PUB1	0.959	0.026	-0.163	0.628	0.018	-0.161
PUB2	0.907	-0.010	-0.351	0.590	-0.007	-0.344
ANTH	0.607	-0.282	0.736	0.448	-0.223	0.818
NUMSEC	-0.060	-0.798	-0.052	-0.024	-0.338	-0.031
ANGSEC	-0.403	-0.814	0.114	-0.190	-0.410	0.081

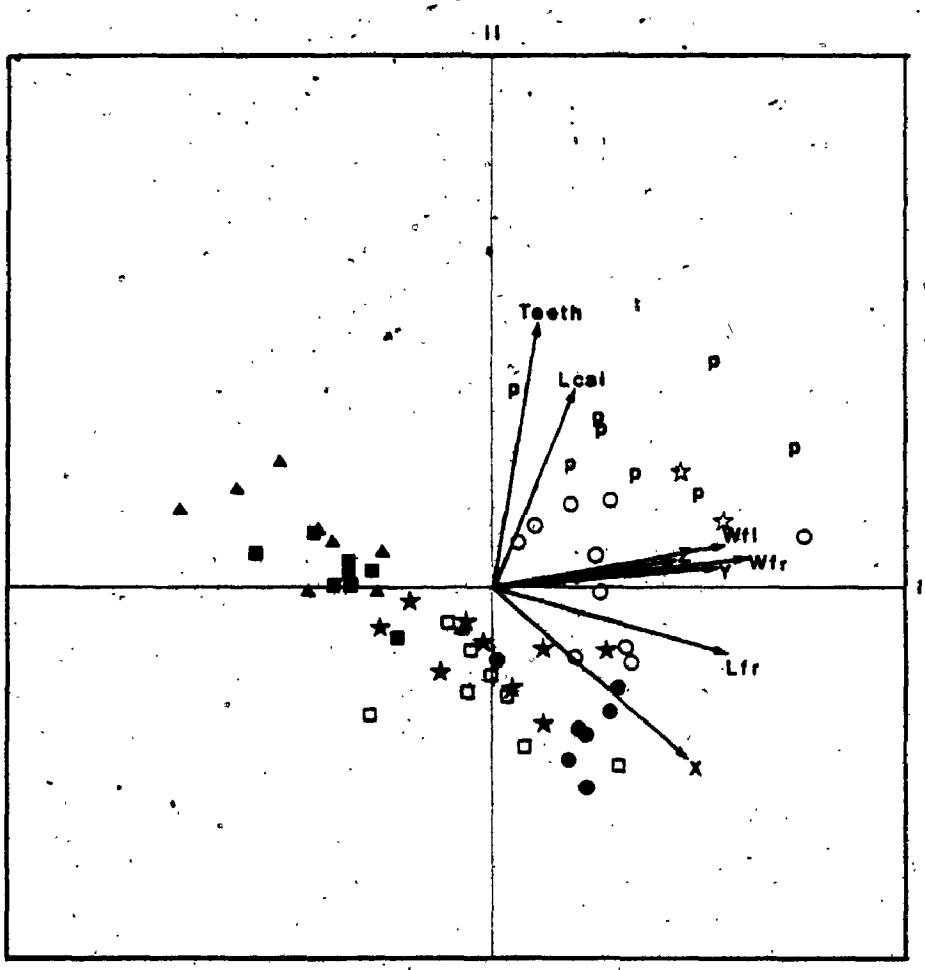


Figure 6.8 60 OTU subsample in the space defined by the first two principal components of the variance-covariance matrix based on 8 ranged continuous flower, fruit and leaf descriptors. Vectors represent the relative magnitudes of component-descriptor correlations in this space. See text and Table 6.6 for details. See Table 3.1 and 3.3 for explanation of descriptor abbreviations. Symbols as in Figure 6.4.

Table 6.6 Parameters of the first three components of the R-PCA of the 60 OTU subsample based on the covariance matrix for 8 ranged continuous flower, fruit and leaf descriptors (see Table 3.1, 3.3 for explanation of descriptor abbreviations).

	Components							
	I	II	III	IV	V	VI	VII	VIII
Eigenvalue	0.17	0.09	0.04	0.03	0.02	0.01	0.01	0.01
% Trace	44.4	22.6	10.8	9.1	5.5	3.3	2.3	1.9
Cumulative %		67.0	77.8	86.9	92.4	95.7	98.0	99.9

	I	II	III	I	II	III
Descriptor	Correlations			Eigenvector Elements		
WFL	0.783	0.141	0.054	0.454	0.115	0.063
LCAL	0.281	0.667	-0.056	0.146	0.486	-0.059
LFR	0.800	-0.234	-0.435	0.439	-0.180	-0.484
WFR	0.866	0.101	-0.363	0.505	0.082	-0.430
X	0.661	-0.584	0.335	0.378	-0.467	0.388
Y	0.756	0.089	0.391	0.326	0.054	0.342
Z	0.616	0.109	0.621	0.262	0.065	0.536
TEETH	0.145	0.895	0.124	0.081	0.697	0.140

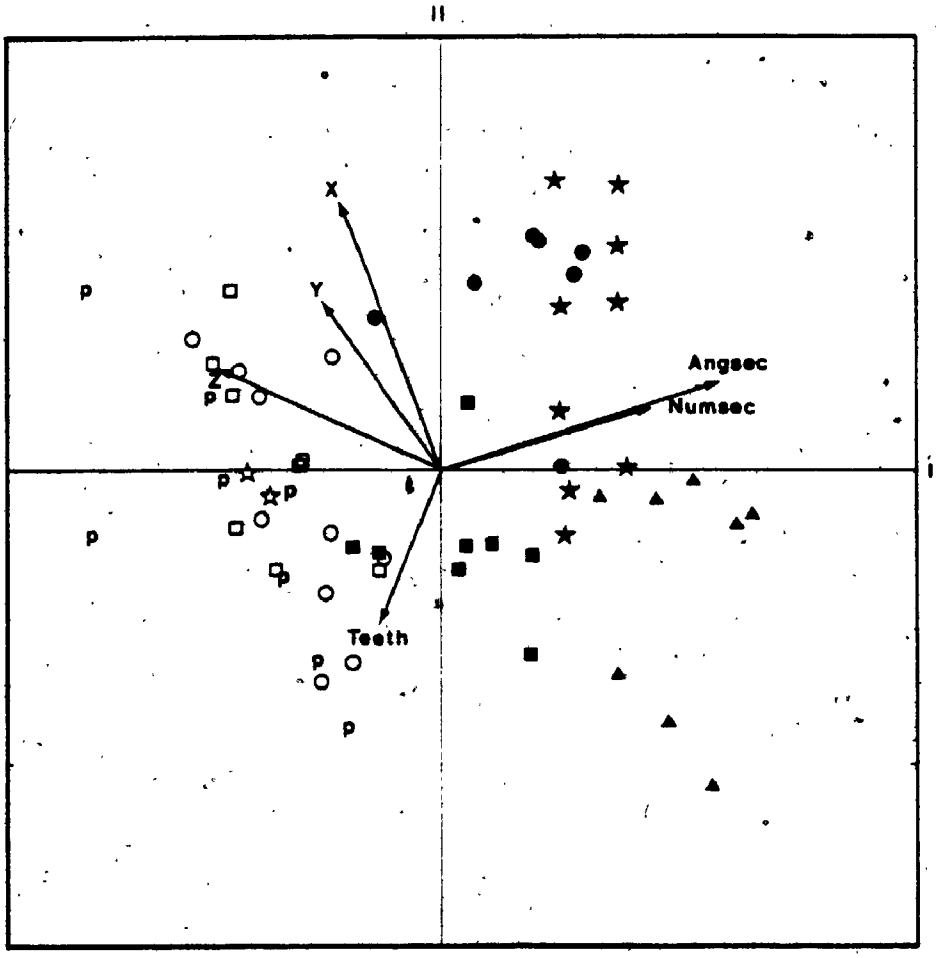


Figure 6.9 60 OTU subsample in the space defined by the first two principal components of the variance-covariance matrix based on 6 ranged short shoot terminal leaf descriptors. Vectors represent the relative magnitudes of component-descriptor correlations in this space. See text and Table 6.6 for details. See Table 3.1 and 3.3 for explanation of descriptor abbreviations. Symbols as in Figure 6.4.

Table 6.7 Parameters of the first three components of the R-PCA of the 60 OTU subsample based on the covariance matrix for 6 ranged short shoot terminal leaf descriptors (see Table 3.1, 3.3 for explanation of descriptor abbreviations).

	Components					
	I	II	III	IV	V	VI
Eigenvalue	0.109	0.080	0.064	0.013	0.009	0.006
% Trace	38.7	28.4	22.7	4.7	3.3	2.2
Cumulative %		67.1	89.8	94.5	97.8	100.0

	I	II	III	I	II	III
Descriptor	Correlations			Eigenvector Elements		
X	-0.341	0.904	-0.010	-0.242	0.748	-0.010
Y	-0.341	0.570	0.618	-0.182	0.356	0.432
Z	-0.747	0.339	0.359	-0.394	0.209	0.248
NUMSEC	0.707	0.210	0.565	0.463	0.161	0.484
ANGSEC	0.921	0.295	0.075	0.719	0.269	0.076
TEETH	-0.213	-0.514	0.794	-0.147	-0.414	0.715

6.3 Effect of Data Set, Resemblance Measure, and Sorting Algorithm on Group Structure

As described above and in Table 4.2, two pairs of contrasting subsets of descriptors were available for the 60 OTU subsample in addition to the complete set of all 17 descriptors. Resemblances between OTUs were calculated for each of these five data sets using two coefficients: euclidean distance (Section 4.2) and generalized distance (Section 4.3.1). An eleventh resemblance matrix was obtained by calculating the information radii (Section 4.4) between OTUs for the 60 OTU subsample. These resemblance matrices were then subjected to three of the sorting algorithms described in Section 5.2. These were single-, average- and complete-linkage, carried out using program NCLAS (Podani, 1980). Illustrations of the results were produced using the plotting routine available with CLUSTAN-1C (Wishart, 1978). Of the 33 phenograms produced only six are reproduced here since with the smaller subsample formal multivariate comparison of all 33 is possible instead, by means of methods developed by Podani (1982).

In contrast with the phenograms in Chapter 5 which are all based exclusively on the same set of eleven flower and fruit descriptors, the 33 phenogams of the 60 OTU subsample were considerably more variable in the extent to which they

reproduced the topodeme structure of the sample (Table 3.2). The three patterns (A-C; Section 5.3) of hierarchical relationships among groups of OTUs described in Section 5.3 were found here as well. Figures 6.10 through 6.15 illustrate some of the ways in which different combinations of data set, resemblance function and sorting algorithm interacted to produce different kinds of cluster composition and differentiation as well as these three patterns and their intermediates.

Figure 6.10 Phenogram of the 60-OTU subsample produced by complete-linkage clustering, using the information radii between OTUs (reference matrix INF). These are calculated on the basis of all 11 flower and fruit descriptors (Table 3.1). Scale at the left represents the minimum resemblance level (information radius) at which the OTUs in any given group were joined by the algorithm used. Symbols as in Figure 6.4. Relationship among clusters represents pattern B (Section 5.3).

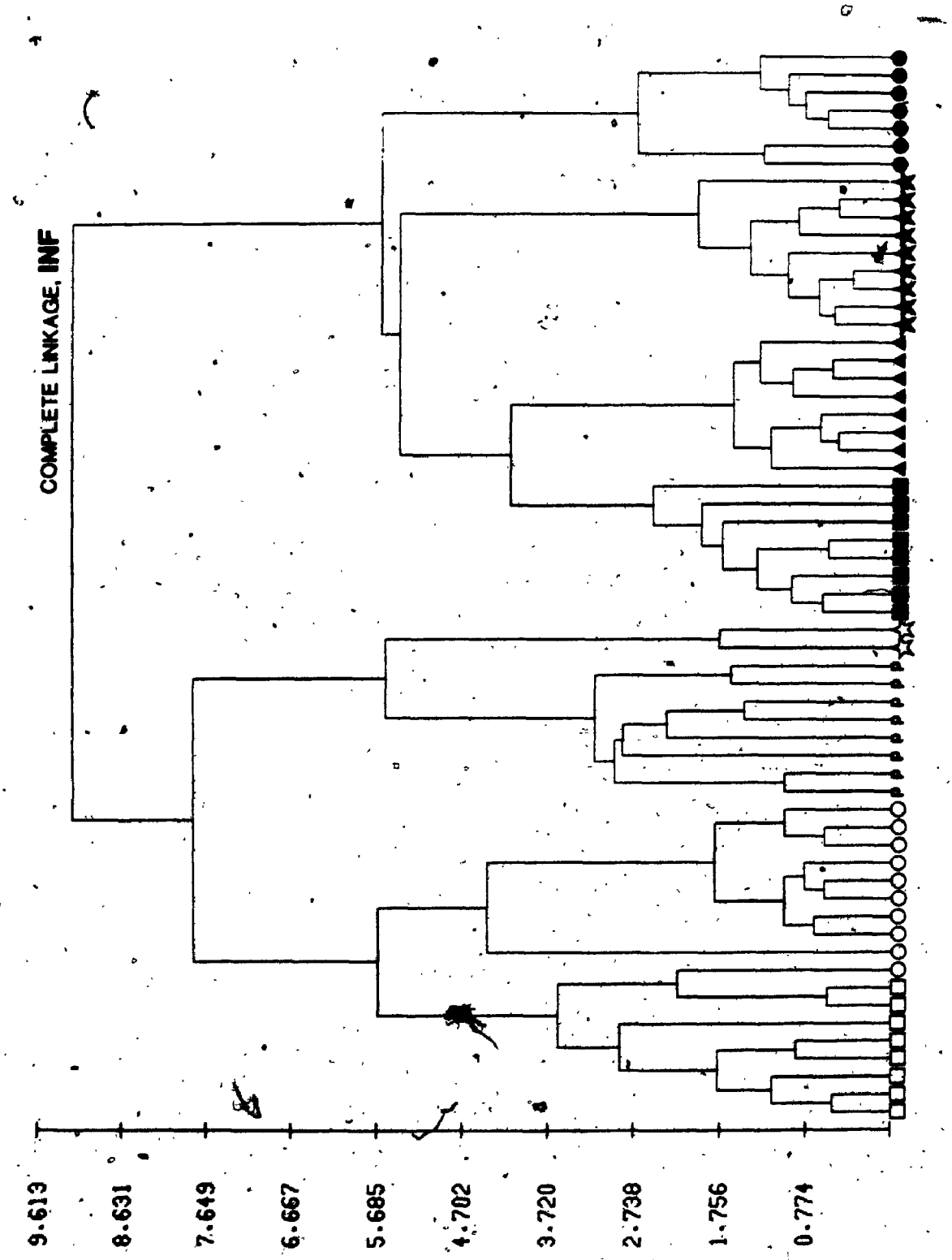

COMPLETE LINKAGE, INF
9.613
8.631
7.649
6.667
5.685
4.702
3.720
2.738
1.756
0.774

Figure 6.11 Phenogram of the 60 OTU subsample produced by average-linkage clustering, using the euclidean distances between OTUs calculated from 9 ranged flower and fruit descriptors (ERE; Table 4.2). Scale at the left represents the minimum resemblance level (euclidean distance) at which the OTUs in any given group were joined by the algorithm used. Symbols as in Figure 6.4. Relationship among clusters represents pattern A (Section 5.3).

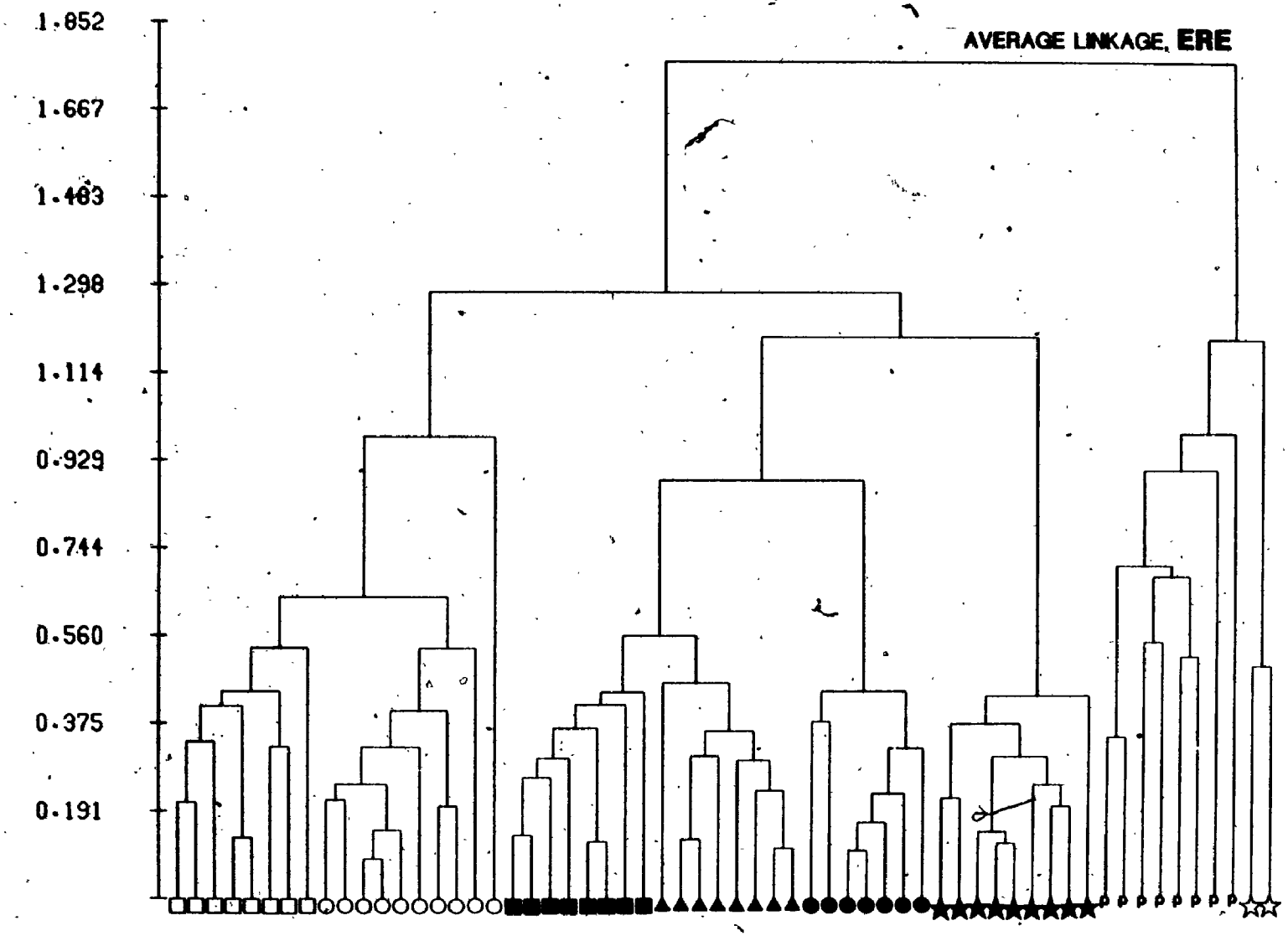

AVERAGE LINKAGE, ERE
1·852
1·667
1·483
1·298
1·114
0·929
0·744
0·560
0·375
0·191

Figure 6.12 Phenogram of the 60 OTU subsample produced by single-linkage clustering, using the euclidean distances among OTUs calculated from all 17 ranged descriptors (EAL; Table 4.2). Scale at the left represents the minimum resemblance level (euclidean distance) at which the OTUs in any given group were joined by the algorithm used. Symbols as in Figure 6.4. Relationship among clusters represents pattern C (Section 5.3).

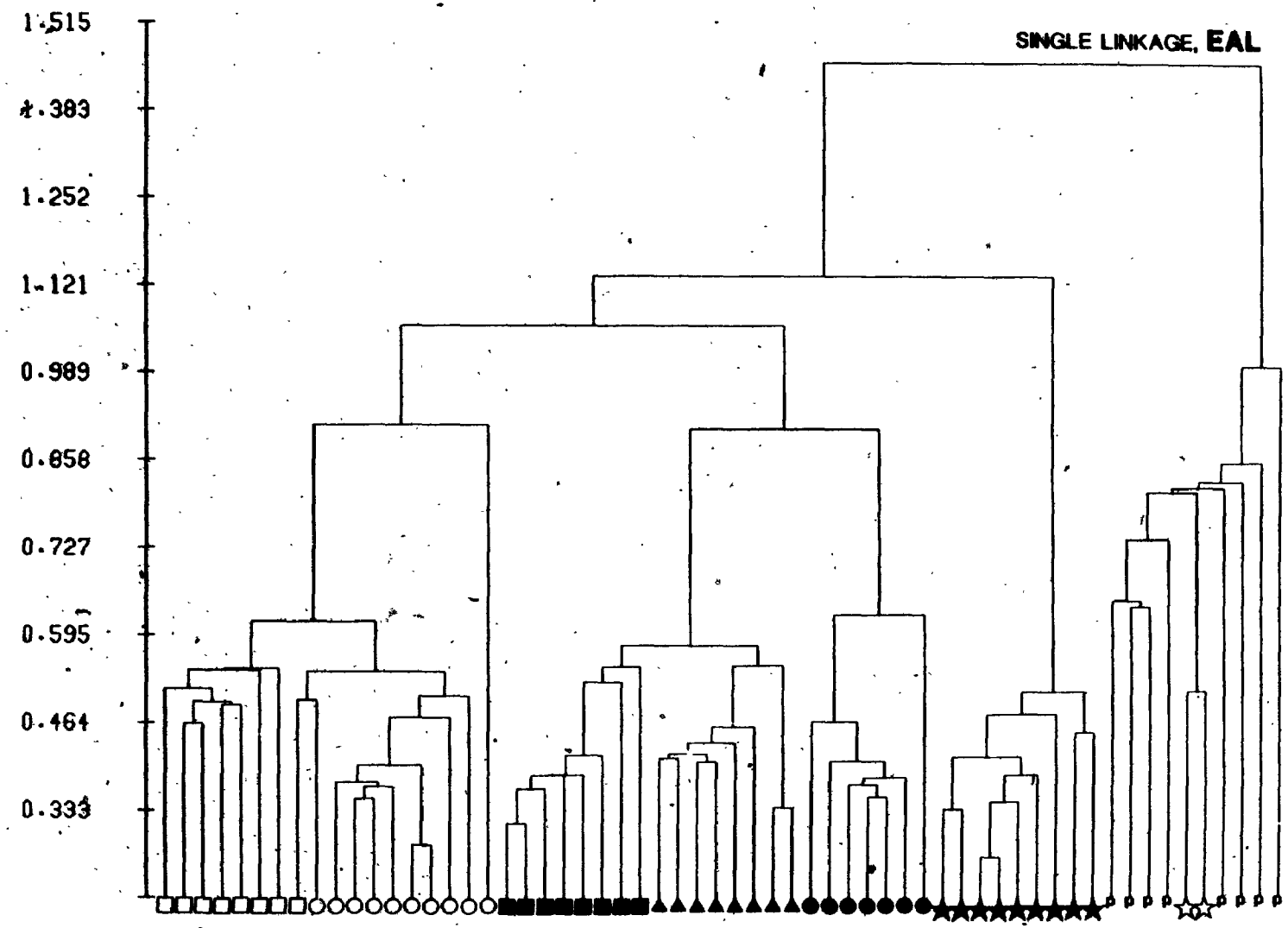

SINGLE LINKAGE, EAL
1·515
1·383
1·252
1·121
0·989
0·858
0·727
0·595
0·464
0·333

Figure 6.13 Phenogram of the 60 OTU subsample produced by average-linkage clustering, using the generalized distances between OTUs calculated from 9 ranged flower and fruit descriptors (GRE; Table 4.2). Scale at the left represents the minimum resemblance level (generalized distance) at which OTUs in any given group were joined by algorithm used. Symbols as in Figure 6.4. The relationship among clusters is intermediate between patterns A and C (Section 5.3).

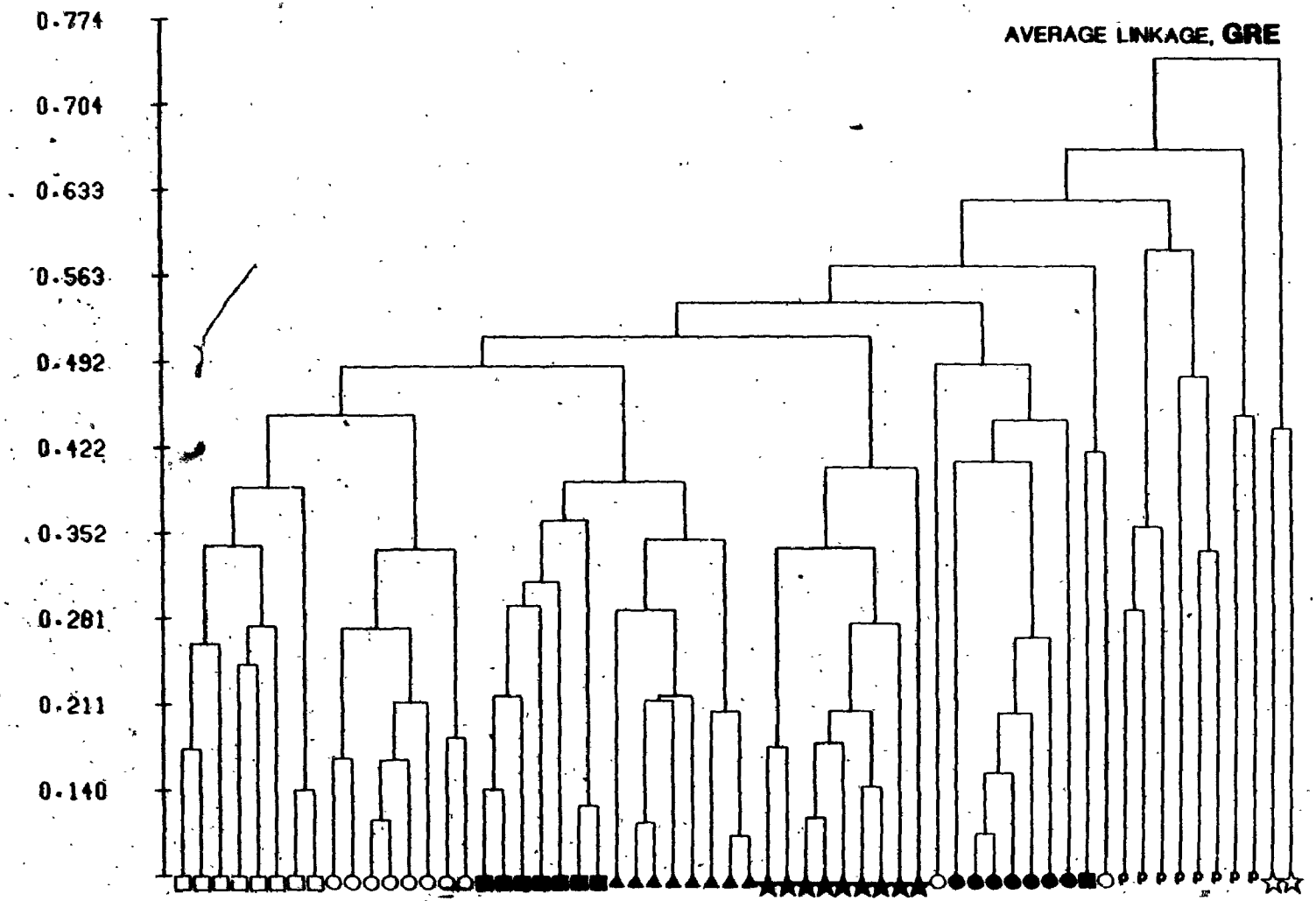

AVERAGE LINKAGE, GRE
0·774
0·704
0·633
0·563
0·492
0·422
0·352
0·281
0·211
0·140

Figure 6.14 Phenogram of the 60 OTU subsample produced by average-linkage clustering, using the euclidean distances among OTUs calculated from 8 ranged continuous descriptors (ECO; Table 4.2). Scale at the left represents the minimum resemblance level (euclidean distance) at which the OTUs in any given group were joined by the algorithm used. Symbols as in Figure 6.4. Relationship among clusters represents a pattern intermediate between A and B (Section 5.3).

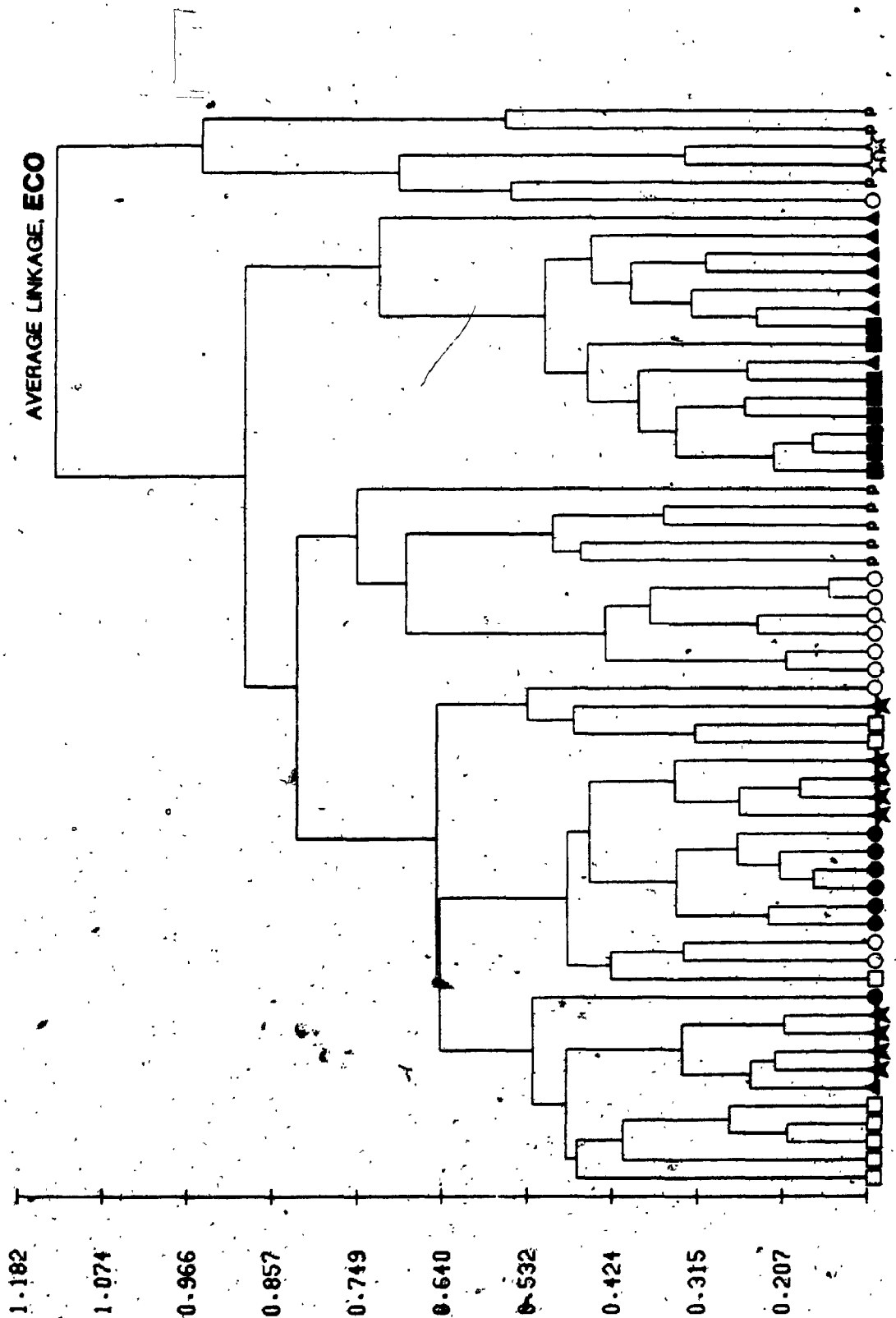
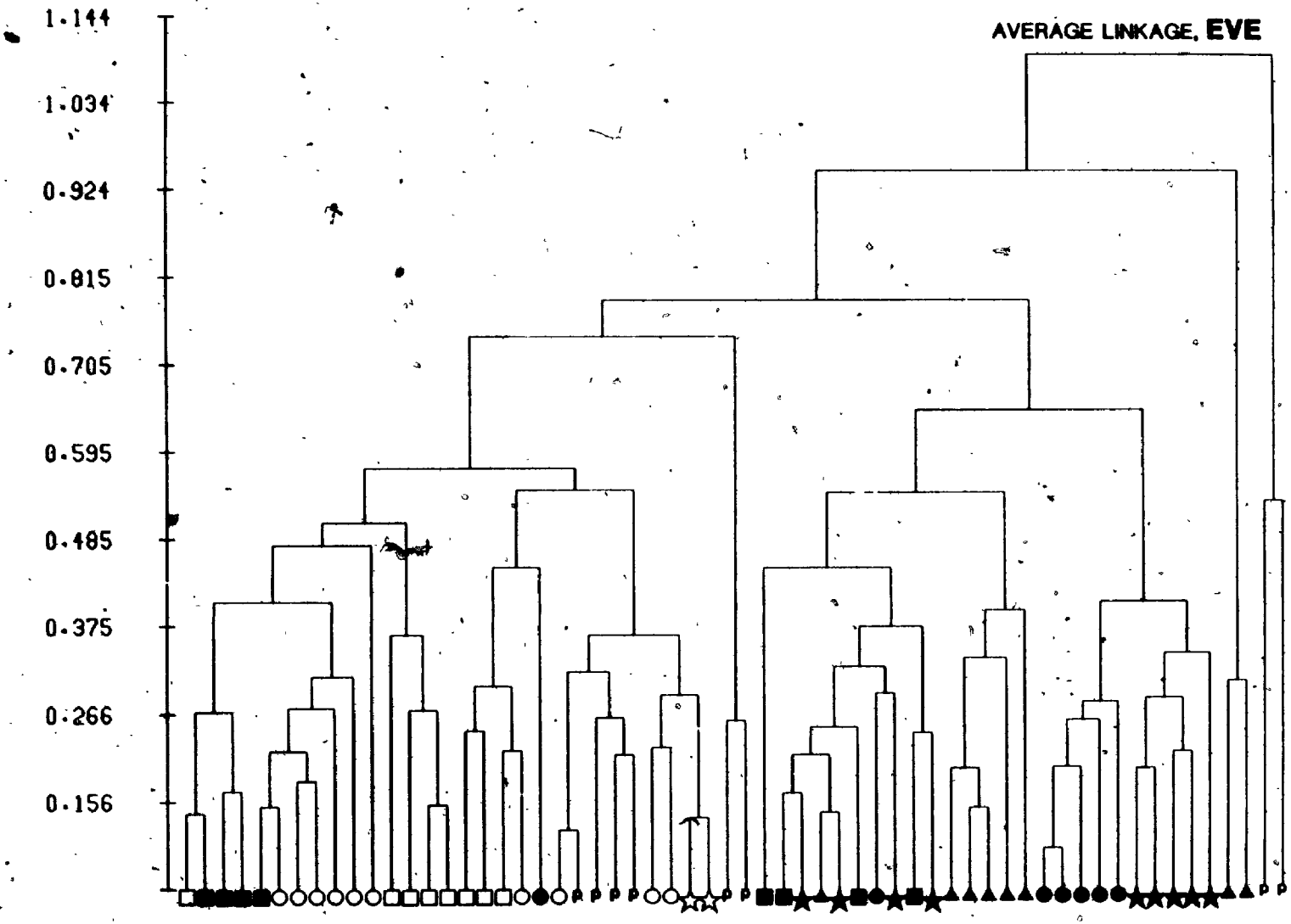

AVERAGE LINKAGE, ECO
1.182
1.074
0.966
0.857
0.749
0.640
0.532
0.424
0.315
0.207

Figure 6.15 Phenogram of the 60 OTU subsample produced by average-linkage clustering, using the euclidean distances among OTUs calculated from 6 ranged short shoot terminal leaf descriptors (EVE; Table 4.2). Scale at the left represents the minimum resemblance level (euclidean distance) at which the OTUs in any given group were joined by the algorithm used. Symbols as in Figure 6.4. Relationship among clusters represents pattern B (Section 5.3).


AVERAGE LINKAGE, EVE
1·144
1·034
0·924
0·815
0·705
0·595
0·485
0·375
0·266
0·156

6.3.1 Multivariate Comparisons of Phenograms; Method

Podani (1982) reviews five descriptors by means of which phenograms may be characterized, and thereby compared. Comparisons of all 33 phenograms based on all five of these descriptors simultaneously have provided a comprehensive picture of the effect of the interaction of data set, resemblance function and sorting algorithm on the results of cluster analyses.

Each of these five descriptors describes a single phenogram by means of a symmetric n x n matrix of pairwise relationships among the n objects in the phenogram. The first is the Cophenetic Value (CV, Sokal & Rohlf, 1962; Sneath & Sokal, 1973). This is the minimum distance between OTUs implied by a particular phenogram.

The second descriptor was suggested independently by Farris (1969), Phipps (1971), and Williams and Clifford (1971), and is referred to here as the Topological Difference (TD). For every pair of OTUs this descriptor takes as its value the number of vertices counted along the path through the phenogram from one OTU to the other.

The remaining three descriptors were developed by Podani himself (1982). These are Cluster Membership Divergence (CMD); Partition Membership Divergence (PMD); and Subtree Membership Divergence (SMD). For each pair of

OTUs the CMD takes as its value the number of OTUs in the smallest cluster containing both OTUs in question. Podani suggests that the effect of this descriptor is intermediate between that of CV and TD. The second descriptor, PMD, takes as its value for any two OTUs the number of partitions of the sample implied by the phenogram in which they are found in different groups. Thus, "PMD preserves the ordering of hierarchic levels although it is a completely topological descriptor" (Podani, 1982). The SMD is related to the PMD, but is expressed, for any pair of OTUs, in terms of the number of subtrees (corresponding to the n - 1 interior vertices of the phenogram) in which they do not co-occur. This last descriptor is unique in that the elements of the principal diagonal of the matrix of values are not all necessarily the same, if the number of interior vertices between an OTU and the root (ultimate fusion) of the phenogram varies (Podani, 1982).

Accordingly, for 5 descriptors of m = 33 phenograms of n = 60 OTUs each, a data matrix (5 x n(n - 1)/2) x m was assembled. Subject to standardization of each descriptor by its global maximum the euclidean distance between phenograms j and k for the n OTUs was obtained as

$$e(j,k) = \left[\sum_{hia}\sum\sum(x(a)_{jhi} - x(a)_{khi})^2\right]^{\frac{1}{2}}, \text{ for } a = 1...5,$$

$$h = 1...n - 1,$$

$$i = h + 1...n,$$

where each value of (a) defines one of the five descriptors CV, TD, CMD, PMD, SMD (program DENDAT; Podani, 1982). The matrix of euclidean distances among the 33 phenograms was then subjected to classification (minimum variance) and ordination (PCoA).

6.3.2 Multivariate Comparisons of Phenograms; Classification Results

Four distinct groups of phenograms (I - IV) could be recognized in the classification results (Fig. 6.16) Group I consisted of three subgroups. Each of these subgroups comprised phenograms conforming to one of the three patterns, A-C (Section 5.3). Pattern A was produced by single- and average-linkage phenograms based on **INF**, plus the average- and complete linkage ones for **EAL**, **ERE** and **EMU** (Fig. 6.11). Only the complete-linkage phenogram produced from **INF** yielded pattern B (Fig. 6.10). The third subgroup is made up of single linkage phenograms in which the OTUs of topodeme sample T6 form a distinct group intermediate between the other OTUs of *C. crus-galli* (both 10- and 20-stamen) and those of *C. punctata* (pattern C, Fig. 6.12).

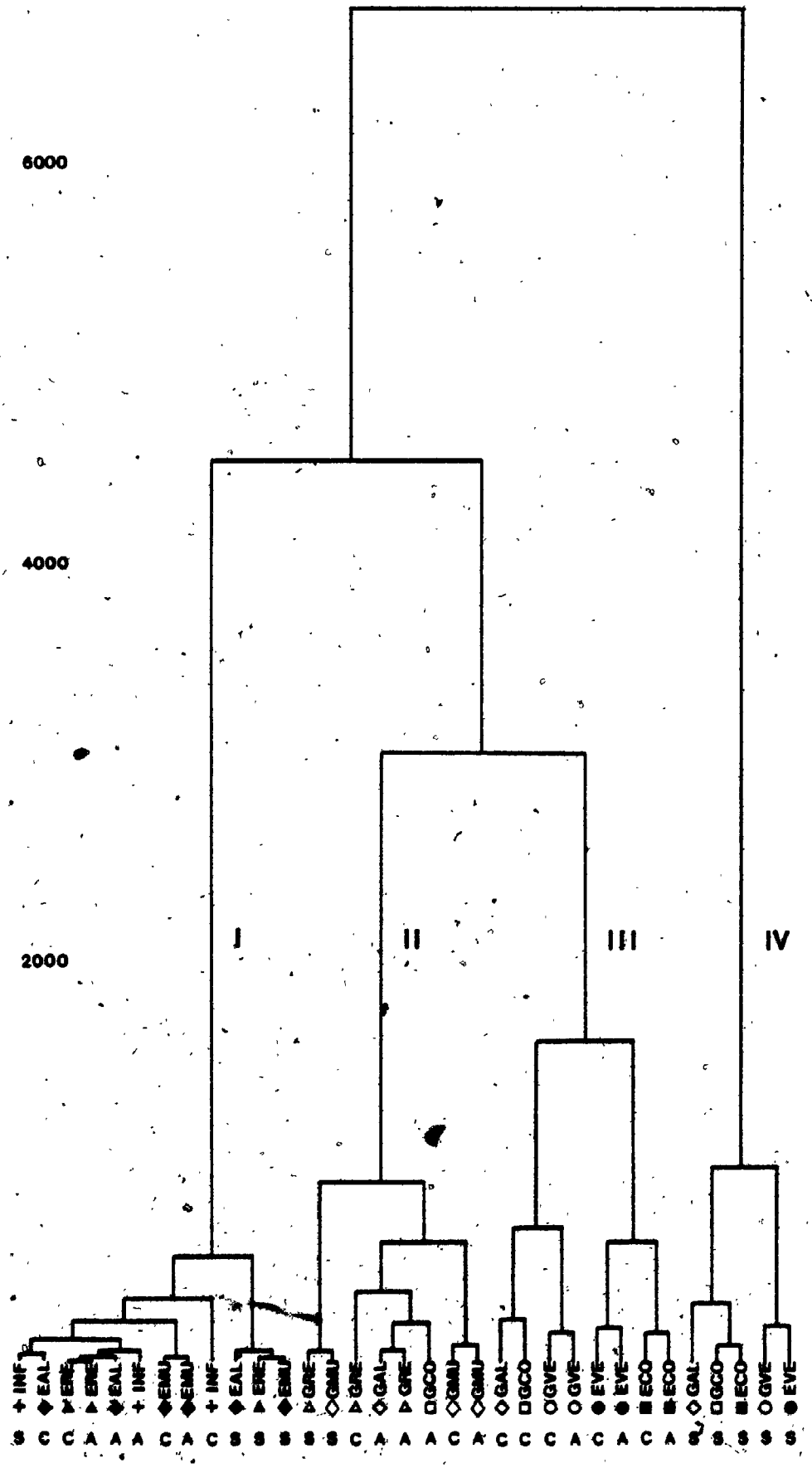
In group II four phenograms (**GRE** and **GMU** with single-linkage, **GRE** and **GMU** with average-linkage; Fig. 6.13) also produce pattern C. The phenograms produced from **GRE** and **GMU** by complete-linkage, and from **GAL** by

average-linkage show or approach pattern A. The remaining phenogram, based on generalized distances calculated from the continuous descriptors (GCO), could not be assigned to one of patterns A-C since although most of the topodeme samples remain largely intact, the two crus-galli morphotypes were not contrasted.

The phenograms in group III produce both pattern A and pattern B, in both cases with some mixing of the topodeme samples and morphotypes. Pattern B is shown by the average-linkage phenogram for EVE (Fig. 6.15) and the complete-linkage ones for ECO and GVE. While the corresponding complete-linkage phenogram of EVE approaches pattern A, the average-linkage ones of ECO and GRE are not categorizable as the crus-galli morphotypes are not clearly distinguished (Fig. 6.14). The complete-linkage phenograms for GAL and GCO approach pattern A.

The fourth group consists of the remaining five single linkage phenograms, for both euclidean and generalized distances based on leaf and continuous descriptors (EVE, ECO, GVE, GCO), plus the one for generalized distances based on all 17 descriptors (GAL). Here, chaining was conspicuous, as is often characteristic of single-linkage clustering, particularly where group definition is weak.

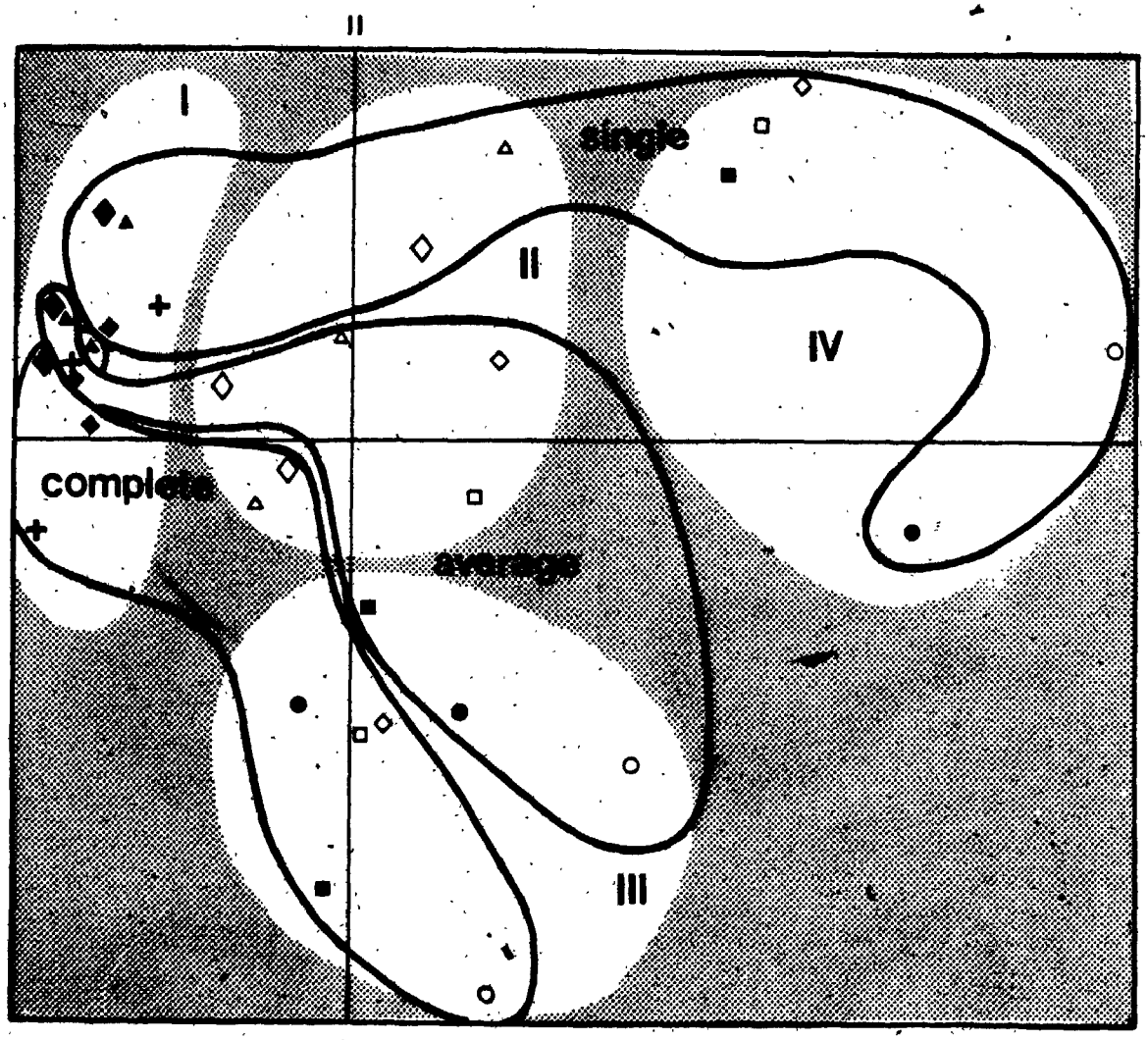
Figure 6.16 Minimum variance classification of 33 phenograms of the 60 OTU subsample, based on the euclidean distances among them (see text for details). Sorting algorithm indicated by letters: S, single-linkage; A, average-linkage; C, complete-linkage. Symbols correspond to the resemblance matrix indicated (see Table 4.2 for explanation of abbreviations), from which each phenogram was calculated: solid symbols, euclidean distance matrices; open symbols, generalized distance matrices; and cross, the reference matrix of information radii (INF). Scale at the left represents the within-group sum of squares for each new group of phenograms.


6000
4000
2000
I
II
III
IV
+ INF
◆ EAL
▶ ERE
▲ ERE
◆ EAL
+ INF
◆ EMU
◆ EMU
+ INF
◆ EAL
▲ ERE
◆ EMU
▷ GRE
◇ GMU
△ GRE
◇ GAL
▷ GRE
□ GCO
◇ GMU
◇ GMU
◇ GAL
□ GCO
○ GVE
○ GVE
● EVE
● EVE
■ ECO
■ ECO
◇ GAL
□ GCO
■ ECO
○ GVE
● EVE
S C C A A A C A C S S S S S C A A A C A C C C A C A C A S S S S S

6.3.3 Multivariate Comparisons of Phenograms; Ordination Results

Although the first two axes of the PCoA of the 33 phenograms account for only 41.7% of the variation in the sample, the ordination in Figure 6.17 illustrates a number of additional features not referred to above. Along the first axis, principally, there is a trend from phenograms based on flower and fruit data and euclidean distances at the left to those based on leaf data and generalized distances toward the right. A second trend, along the second axis, is one from phenograms produced by the space-contracting algorithm (single-linkage) through those produced by average-linkage to ones produced by the space-dilating algorithm (complete-linkage). Superimposition of outlines of groups I - IV on the ordination shows the intermediacy of the phenograms in groups II and III with respect to those in group I and group IV (compare Fig. 6.16 and 6.17).

Figure 6.17 33 phenograms of the 60 OTU subsample in the space defined by the first two principal coordinate axes of the matrix of euclidean distances among them. See text for details. Sorting algorithm used (single-, average- and complete-linkage) as indicated by the solid closed lines. Unstippled areas indicate the four clusters of phenograms (I - IV) obtained in Figure 6.16. Symbols indicate the data sets and resemblance functions used to obtain the corresponding phenograms, and are explained in Figure 6.16 by reference to Table 4.2: all 17 descriptors, short lozenges; 9 flower and fruit descriptors, triangles; 7 multistate descriptors, long lozenges; 8 continuous descriptors, squares; and 6 leaf descriptors, circles. Solid symbols: phenograms calculated from matrices of euclidean distances; open symbols, phenograms calculated from matrices of generalized distances. Phenograms calculated from the reference matrix of information radii (INF) are indicated by crosses.


II
I
single
II
IV
I
complete
average
III

6.4 Discussion

In Chapter 5 cluster analyses of the entire 160 OTU sample resulted in conspicuous evidence of group structure in the flower and fruit data. The groups found corresponded largely, in the first instance, to individual topodeme samples, and in the second, to the five or six taxa distinguishable in the sample: the two common morphotypes of C. crus-galli; C. sp. aff. C. bushii; C. ?grandis; C. punctata; and C. Section Macracanthae.

The studies presented here have been concerned principally with understanding how the data obtained distinguish the topodeme samples in the 60 OTU subsample. On one hand, these groups of OTUs can be related to particular univariate contrasts and to subjective impressions from the field. On the other hand, as discussed in the following two chapters, they can be subjected to multivariate tests of their validity. Here, however, it is important first to understand the extent to which subjective impressions and univariate contrasts have been confirmed merely by features of data type and method of analysis rather than by the data itself.

6.4.1 Data Type

The descriptors of reproductive and vegetative morphology used in this study (Chapter 3) are a mixture of binary, multistate and continuous ones. The reason for this is that features that are useful in distinguishing taxa are often contrasts between states, or based on comparisons of size or shape. Although those used here have been ranged so as to make them both commensurate and equally weighted (Section 4.5.2), it was found in supplementary analyses described in Section 5.4.4 and 6.2.1 that the descriptors still vary considerably in their distributions. As a result multistate and continuous descriptors (Table 4.2) make quite different contributions to the group structure perceived in the sample by means of either classifications (Chapter 5, Section 6.3) or ordinations (Secton 6.2).

This contrast is clear also from a comparison of Fig. 6.7 and Fig. 6.8. In the former, 7 multistate descriptors distinguish topodeme samples more than do 11 flower and fruit descriptors, including style number and stamen number, in Figure 6.4. In Figure 6.8 (continuous descriptors), however, the topodeme samples intermingle considerably, and the corresponding hierarchical structure imposed by classification (Fig. 6.14) is difficult to interpret.

A similar contrast is seen in a comparison of ordinations based on leaf data (Fig. 6.9) with those based on the flower and fruit data (Fig. 6.4, 6.8). While individual topodeme samples are to be distinguished by the leaf descriptors as more or less discrete groups of OTUs, a greater discontinuity appears to exist between a group of OTUs in which stamen number is characteristically high (20-stamen crus-galli and C. punctata) and the 10-stamen crus-galli (compare Fig. 6.15). Comparison of the classifications based on these two data sets indicates that if resemblance function and sorting algorithm are held constant (Fig. 6.11, 6.15) the contrast between them relates to the resolution of the topodeme samples by the clusters formed.

6.4.2 Resemblance Function

This section deals with comparisons of the 33 classifications of the 60 OTU subsample discussed above (Section 6.3). The separation observed in Figures 6.16 and 6.17 of phenograms based on the matrix of generalized distances for reproductive descriptors (GRE) from those based on the reference matrix of information radii, for a given sorting algorithm, was unexpected in view of the formal relationship between these two resemblance functions described by the originators of the latter (Sibson, 1969;

Jardine, 1971; Jardine & Sibson, 1971). However, these results may be due to failure to sufficiently meet the requirement that the descriptors used to calculate the generalized distance have a joint, multivariate normal distribution. Both information radii and euclidean distances are insensitive to this circumstance.

Comparing the position in Figure 6.17 of the average linkage phenograms based on **ERE** and **GRE** (solid and open triangles, respectively), and the appearance of these two phenograms in Figures 6.11 and 6.13 it appears that the displacement to the right along the first ordination axis corresponds in part to a shift from pattern A (**ERE**) to one intermediate between A and C (**GRE**). It also corresponds to a decrease in the sharpness of group structure. This is expressed as the contrast between the levels at which clusters are formed and those at which they are joined to other clusters.

In general, clusters in the phenograms obtained from matrices of generalized distances are less differentiated from one another than is the case in the corresponding phenograms based on euclidean distances (compare Fig. 6.11 and 6.13). Again, this recalls similar contrasts seen in Chapter 5. One reason for the resulting dissociation of analyses based on generalized distances from the corresponding ones based on euclidean distances (Fig. 6.16,

6.17) is the way in which the latter preserve the effect of discriptor correlations. Classifications of the sample based on generalized distances represent the group structure that exists independent of descriptor correlations. The contrast between euclidean distance based classifications and those based on the generalized distance seen here and in Chapter 5 suggests that the group structure present in the samples studied exists independently of, but is strongly reinforced by the correlations among descriptors.

Finally, the similarity of the classification and ordination results based on information radii among both 60 and 160 OTUs to those of similar analyses based on euclidean distances deserves further comment. The information radius was employed here because of its independence from assumptions concerning the amount and pattern of within-OTU variability in the descriptors used. The similarity of results calculated from descriptor means to those based on information radii calculated from observed descriptor-state frequencies (compare Fig. 6.4 and 6.6; 6.10 and 6.12, as well as the phenograms in Chapter 5). suggest that descriptor means provided an adequate representation of descriptor expression in individual OTUs.

6.4.3 Sorting Algorithm

The trend related to sorting algorithm along the second PCoA axis in Figure 6.17 has already been noted in Section 6.3.3 above. This trend reflects differences in the hierarchical relationships among clusters, from pattern C phenograms at the top of the figure to pattern A and then to pattern B, at the bottom. At one extreme (pattern C), subtle differences between clusters are preserved by the space-contracting effect of single-linkage clustering. At the opposite extreme, the space-dilating effect of complete-linkage clustering operates to emphasize gross differences, as in stamen number, at the expense of finer ones (pattern B).

6.4.4 Interaction between Data Set, Resemblance Function, and Sorting Algorithm

As described above, the first two PCoA axes in Figure 6.17 appear to represent trends in two features of phenogram structure: cluster definition, in terms of both composition and resolution, and the hierarchical relationships among clusters, respectively. It was evident from the asymmetry of the distribution of classifications around these two axes that the factors discussed above, data set, resemblance function, and sorting algorithm did not act independently of one another. To illustrate, consider the displacements along the first axis of the phenograms based on matrices of

generalized distances with respect to those based on the corresponding matrices of euclidean distances. As described in Section 6.4.2 above, this displacement is related to the contrast between the level at which a cluster may be recognized and that at which it joins another cluster at a higher level. For the single-linkage phenograms these displacements are much greater than are those for the other two algorithms. This is due to the sensitivity of single-linkage to the presence of intermediate OTUs, such that they are joined to and connect existing clusters rather than becoming the nuclei of new ones. As a result, the clusters built up in this way tend to be poorly differentiated from one another. The positions of the single-linkage phenograms in Figure 6.17 further suggest that a similar interaction has taken place between sorting algorithm and data type.

An implication of the interaction between sorting algorithm and resemblance function is that cluster definition in the sample here is to a considerabe extent a product of descriptor correlations, in view of the effect of removal of those correlations on phenogram structure.

6.4.5 Conclusions

The analyses presented here document the marked effect the data set, resemblance function, sorting algorithm, and their interactions may have on the relationships which may

be perceived among a sample of OTUs.

In both the entire 160 OTU sample and the 60 OTU subsample, however, there is at the same time a striking consistency in the way in which the groups detected consisted predominantly of OTUs from the same topodeme sample. Where this was not the case appears to be due to interaction between sorting algorithm, resemblance function and the data sets in which group structure was the weakest and one sorting algorithm (single linkage).

There is however no suitable way of choosing among the alternative classifications (e.g. patterns A, B, and C; Fig. 6.16, 6.17) a single best, or group of best classifications. As discussed above, each group of classifications (I-IV) and each pattern of hierarchical relationships (A-C) is to some extent a function of one data type as opposed to another, or one sorting algorithm as opposed to another. Some kind of extrinsic information is required in order to evaluate these alternatives. In the following two chapters methods are employed which enable an evaluation of the success with which the data for the entire random sample support a group structure based on the topodeme sample affiliation of the randomly sampled OTUs. Chapters 9 and 10 then present independent evidence that provides a possible biological model (Chapter 11) for such a group structure.

The choice among the three patterns of hierarchical relationships implied by the data is less easy to test, but at the same time also easier to resolve. Testing requires a larger and geographically more extensive sample in order to determine, for example, whether the form represented by topodeme sample T6 (Site 4) is anything more than an extremely distinct but merely local variant. It is also necessary to determine to what extent this form is to be distinguished on the basis of features other than anther color.

Likewise, other lines of evidence (cytological, phytochemical) are required to determine whether the intermediacy of 20-stamen C. crus-galli with respect to 10-stamen C. crus-galli and C. punctata is anything more than a function of the selection of descriptors and a number of independently evolved shared descriptor states. The remaining pattern (A) is considered to represent the simplest alternative: that within the scope of the comparison made here, C. punctata is an outgroup equally to both of the common morphotypes of C. crus-galli.

Resolution of this situation in the absence of additional evidence is possible by accepting all three relationships (A-C) as equally valid expressions of the information present in the data available.

CRATAEGUS CRUS-GALLI L. SENSU LATO IN SOUTHERN ONTARIO: Phenotypic Variation and Variability in relation to Reproductive Behavior

by

Timothy Adam DICKINSON

Department of Plant Sciences

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario

June, 1983

VOLUME TWO

TABLE OF CONTENTS

(Volume 2)

Page

* * *

CHAPTER SEVEN

COMPARISON OF GROUP COVARIANCE STRUCTURES

7.1 Introduction

In the preceding two chapters clustering and ordination methods readily differentiated a number of groups within the 160 OTU sample. With respect to the random sample these groups corresponded to the individual topodeme samples. The present chapter is concerned with comparisons of both the degree and pattern of variation within these samples.

Such comparisons are of interest for a number of reasons. In the first place comparisons of the amount of variation present are a prerequisite for the statistical evaluation of topodeme sample distinctness (Chapter 8). Secondly, as discussed in Chapters 10 and 11, such comparisons bear on questions concerning the results of different modes of reproductive behavior. These in turn have implications for the taxonomic treatment of the groups making up Crataegus crus-galli sensu lato. Thirdly, the patterns of descriptor variation and covariation themselves, within and between groups, also convey information about group structure, but are not often examined directly. Finally, studies of morphogical variability in samples from

plant populations appear to have employed principally univariate comparisons. Although reference is made to univariate contrasts, the present study places greater emphasis on multivariate methods which assess the covariation of descriptors as well as their individual variances. The present study thus offers an opportunity to examine the performance of these methods in connection with the objectives described above.

7.2 Univariate Comparisons of Variability

Univariate comparisons of variability have employed the standard deviation of individual descriptors, often either calculated from logarithmically transformed data (Lewontin, 1966) or rescaled by the corresponding mean to achieve commensurability, resulting in the latter case in a coefficient of variation (CV; Băra, Ghiorgită & Toth, 1973; Băra & Ghiorgită, 1973; Yablokov, 1974; Usberti & Jain, 1978).

Simpson, Roe, and Lewontin (1969) caution that it is important that comparisons of CVs be made across homologous descriptors only, not only in terms of the structures whose variability is being compared but also in terms of data type (e.g. multistate versus continuous). Sokal and Braumann (1980) have shown that CVs are extremely sensitive to departures from normality, and have provided guidelines for

their statistical interpretation. Day and Fisher (1937; also Van Valen, 1978), however, point out that the assumptions underlying such a description of sample variability even with respect to homologous and normally distributed descriptors, namely that variability increases in proportion to increase in the mean are not necessarily confirmed experimentally (Fig. 7.1). They propose instead an analysis of the covariance of the means and their standard deviations in order to adjust the latter for the effect of the magnitude of the mean. This approach has been adopted for morphometric studies of Plantago maritima L. (Gregor & Lang, 1950) and species of Armeria (DC) Willd. (Baker, 1953). The method cannot be applied to the crus-galli-punctata comparison here since only a single topodeme of the latter species has been sampled. However, examination of Figure 7.1 indicates that for the descriptors shown this method would be more appropriate than use of the CV since among the crus-galli topodeme samples the standard deviation does not appear to increase mohotonically with increasing mean.

An alternative to the variance (on which the CV is based) and one which is robust to departures from normality is Levene's statistic $\bar{y}$. This statistic examines the average deviation of variates x_i from their mean in a sample of size n (Van Valen, 1978), where

$$y_i = |x_i - \bar{x}|, \quad i = 1...n.$$

Using analyses of variance or t-tests, Levene's $\bar{y}$ can be compared directly among samples and among commensurate descriptors, or after being rescaled by the mean, if that is appropriate. Van Valen (1978) suggests that the performance of $\bar{y}$ may be improved by substituting for $\bar{x}$ either the modal value of x_i or a robust estimate of the mean obtained from the median 80% of the data (i.e. by excluding 10% of the values at each end of the distribution).

7.3 Multivariate Comparisons of Variability

Warwick and McNeill (1982) found that population samples of Plantago major L. from southern Ontario fell into three main groups. The frequency with which individual samples were misclassified by discriminant analyses was taken as an indication of their variability.

A related approach to the multivariate comparison of variability has been proposed by Gilmartin (1974, 1980), and is based on the resemblances (distances) between OTUs used in cluster analyses. For a given sample, a resemblance matrix is calculated and from this, the mean resemblance value for the sample (Mean Phenetic Distance, MPD). The corresponding standard deviation is also calculated, and used to obtain, by analogy with the CV, the Coefficient of Phenetic Variability (CPV). The method thus enables

comparison of the magnitude of between-OTU variability on the basis of average between-OTU distances (MPD), as well as the variability of that quantity over a number of homologus groups (CPV). A result similar to the MPD may be obtained by direct inspection of phenograms, comparing the resemblance levels at which clusters are formed (Fig. 5.1-5.12; 6.10-6.15).

Gilmartin's method appears to be more widely applicable than the others described here, since it can be used with data sets which include descriptors that are invariant within a given group. However, for the reasons discussed in Section 5.4 and Chapter 6, this method may give misleading results if effects solely of data type and resemblance function are not considered when comparing samples.

Van Valen (1974, 1978) has suggested a third multivariate method for comparing the variability of samples. This is the examination of their total variance, over all p (commensurate) descriptors considered. It leads to a multivariate generalization of Levene's statistic $\bar{y}$ calculated from

$$y_i = \left(\sum_j (x_{ij} - \bar{x}_j)^2\right)^{\frac{1}{2}}, \quad i = 1 \ldots n, \; j = 1 \ldots p.$$

Thus $\bar{y}$ is the average euclidean distance in p-dimensional space between OTUs and the centroid of the group to which they belong. As in the univariate case differences among

samples with respect to $\bar{y}$ may be evaluated by means of analyses of variance or t-tests.

Finally, a completely different means of illustrating the multivariate variability of a sample is the method for two-dimensional summarization of high-dimensional data proposed by Andrews (1972; Gnanadesikan, 1977). The elements of the data vector for each OTU are used as coefficients of the terms of a trigonometric function. As a result, a standing wave is perturbed in a characteristic way by each OTU. The extent to which OTUs in a sample vary is indicated by the width of the band created by their superimposed wave plots. Since when used in this way the method is purely descriptive, rather than quantitative, it is not considered further here, although its use has corroborated the results obtained below.

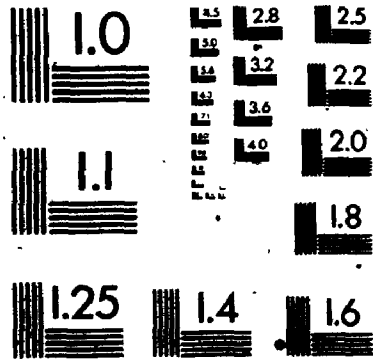
7.3.1 Comparison of Dispersion Matrices

Multivariate variability may also be quantified by reference to the structure of dispersion matrices. Dispersion matrices may be compared informally by examination of their determinants (Goodman, 1968; Mitton, 1978). This scalar is a multivariate generalization of the variance (Wilks, 1932; Cooley & Lohnes, 1971; Pimentel, 1979), and is a unique function of any square matrix (the product of its eigenvalues).

Alternatively, comparisons of sample dispersions may be made by statistically testing null hypotheses of non-differentiation (e.g. Handford, 1980). This is the approach employed here because as mentioned earlier it is a prerequisite for testing the differentiation of topodeme sample centroids in Chapter 8.

The test of the equality of sample dispersion matrices described below is based on comparisons of their determinants. Because of this it is extremely sensitive to features of the data affecting these quantities. In particular, as one or more variances approach zero, or as most of the correlations approach one, the determinant of a dispersion matrix approaches zero (Van Valen, 1974, 1978). As described in Section 7.6 the test is also quite sensitive to departures from multivariate normality. Accordingly it has been supplemented by other methods of comparison in order to have a better basis for interpreting the results obtained. The multivariate $\bar{y}$ described above has been employed as an alternative metric of variability, while the graphical methods described in Section 7.6 have been used to provide a more detailed description of the dispersion matrices being compared.

4



7.3.2 Method

Formal hypotheses concerning group differentiation and the equality of group dispersions are tested in the context of the decomposition of the total sample variance into among-group and pooled within-group components. For g groups each made up of N_k individuals (k = 1...g), the matrix $\underset{\sim}{T}$ of sums of squares and cross-products (SSCP) of deviations of all individuals from the grand centroid for the total sample is described by $\underset{\sim}{T} = \underset{\sim}{A} + \underset{\sim}{W}$, or

$$\sum_{ki}\sum \underset{\sim}{x}_{ki}\underset{\sim}{x}'_{ki} = \sum_{ki}\sum(\underset{\sim}{m}_k - \underset{\sim}{m})(\underset{\sim}{m}_k - \underset{\sim}{m})' + \sum_{ki}\sum(\underset{\sim}{X}_{ki} - \underset{\sim}{m}_k)(\underset{\sim}{X}_{ki} - \underset{\sim}{m}_k)',$$

for k = 1...g, i = 1...N_k, where $\underset{\sim}{m}$ is the grand centroid, $\underset{\sim}{m}_k$ is the centroid of the kth group, $\underset{\sim}{X}_{ki}$ is the data vector for the ith individual of the kth group, and $\underset{\sim}{x}_{ki}$ is the same vector centered by subtraction of the grand mean. The first term on the right represents the SSCP matrix for deviations of the group centroids from the grand centroid ($\underset{\sim}{A}$, for 'among-groups'), while the second represents the SSCP matrix for deviations of individuals from their group centroids, pooled over all groups ($\underset{\sim}{W}$, for 'within-groups').

The null hypothesis that none of the group centroids differ from the common population centroid μ, i.e. that $\underset{\sim}{\mu}_k = \underset{\sim}{\mu}$ for k = 1...g, is tested in Chapter 8. The multivariate analysis of variance which tests this hypothesis assumes the homogeneity of all the group

dispersions. Consequently, a prior null hypothesis of the equality of group dispersions, i.e. that $\underset{\sim}{\Delta}_k = \underset{\sim}{\Delta}$ for k = 1...g, must first be tested.

A criterion for testing the equality of group SSCP matrices is due to Box (1949), based on earlier work by Bartlett (1937). It employs the determinants of the covariance matrices derived from the individual group SSCP matrices ($\underset{\sim}{D}_k$), together with the determinant of the pooled within-group covariance matrix derived from $\underset{\sim}{W}$, here referred to as $\underset{\sim}{D}_w$.

$$\text{Box's } M = (N - g)\ln|\underset{\sim}{D}_w| - \sum_k (N_k - 1)\ln|\underset{\sim}{D}_k|, \quad k = 1...g.$$

The multivariate analysis of variance described above was carried out using program MANOVA (Cooley & Lohnes, 1971), which calculates M as well as its corresponding F value.

7.3.3 Descriptor Selection

The multivariate analysis of variance, and particularly Box's M-test have a formal requirement that the data analyzed have a multivariate normal distribution. In addition, the tests of equality of group dispersions and group means also require that the number of descriptors used not exceed the number of OTUs in the smallest group. In the present case, with the inclusion of the seven OTUs of

topodeme sample T5, it is convenient to restrict the analysis to six descriptors. The two OTUs of C. ?grandis found in the random sample at Site 4 (Table 3.2) were omitted from the analyses described here for this reason. A third requirement is that descriptors not be invariant within any group. Descriptors such as pubescence and anther color (Section 6.2) must be ruled out of the analysis on this basis.

Although examples to the contrary may be found (in Harris, 1975, for instance), it appears advisable also to restrict the analysis here as much as possible to continuously distributed descriptors because of the greater likelihood of their joint distribution approaching multivariate normality, than if binary or multistate descriptors were used. Accordingly, the six flower and fruit descriptors chosen for use here include all four continuous descriptors (WFL, LCAL, LFR, WFR), plus the meristic descriptor STYL, plus one ordered multistate descriptor, TCAL (see Table 3.1 for explanation of descriptors and their abbreviations). An additional factor in this selection was a desire to avoid unnecessarily high descriptor correlations due to descriptor redundancy. Descriptors such as stamen number (STAM) and pubescence (PUB1, PUB2) are highly correlated with a number of other descriptors, including style number (STYL) and flower width

(WFL; Table 3.8, 3.10; Table 8.4). Consequently, inclusion of the latter justified exclusion of the former. In the case of the six leaf descriptors (Table 3.3), selection was not necessary. These six include three continuous descriptors (X, Y, Z), two meristic ones (NUMSEC, TEETH), and one multistate one (ANGSEC).

7.4 Leaf Data Results

In Chapter 6, analyses of the leaf data available for the 60 OTU subsample suggested two possible group structures. These were subdivision of the sample based on stamen number (2 groups) and based on topodeme sample affiliation (7 groups). A third possibility, based on the similarity in leaf shape of sympatric topodeme samples (illustrated in Chapter 8) was subdivision by site (5 groups). Of these, only the hypothesis of seven groups corresponding to the individual topodeme samples resulted in a non-significant value of F (Table 7.1). In each of the other two cases $H_0: \underset{\sim}{\Delta}_k = \underset{\sim}{\Delta}$, k = 1...g is rejected.

Comparison of the seven topodeme sample total variances (Table 7.3a) and dispersion matrix determinants (Table 7.4a) reflected this lack of difference between the groups.

7.5 Flower and Fruit Data Results

The seven group hypothesis was tested for the six flower and fruit descriptors selected in Section 7.3.2 above. This was done with both the entire 111 OTU random sample of the seven topodemes studied and with the 58 OTU subsample of T1 - T7 (Table 3.2). Exclusion of the punctata sample (T7) led to a test of a hypothesis of the homogeneity of covariance matrices of six exclusively crus-galli groups (93 OTUs). All three tests indicated significantly heterogeneous topodeme sample dispersions (Table 7.2). However, it is of interest that the level of significance at which the null hypothesis is rejected decreases from the 111 OTU sample containing both crus-galli and punctata to both the 93 OTU crus-galli sample and the 58 OTU sample also containing punctata (Table 7.2). In the latter case this change may be a function of there being less disparity in size among the topodeme samples (Table 3.2). The former case is more interesting. Since removal of the punctata sample results in greater homogeneity of group dispersions, it suggests that the dispersion of the latter sample in particular differs appreciably from that of the crus-galli samples.

Examination of the total variances ($\bar{y}$) calculated for the flower and fruit data (Table 7.3b) and of the determinants of the covariance matrices for the same

descriptors (Table 7.4b, c) for each topodeme sample indicates that the dispersion of the punctata sample (T7) is in fact appreciably greater than that of any of the crus-galli ones (T1 - T6).

The effect of data type on this hypothesis of a taxonomically based difference in covariance structure was tested by repeating the analyses omitting first TCAL (ordered multistate) and then STYL (meristic). However, in both cases the same pattern emerged with respect to the test of covariance matrix homogeneity (Table 7.2) and with respect to the relative magnitudes of the determinant of the topodeme sample covariance matrices (Table 7.3c).

7.6 Graphical Comparisons of Covariance Matrices

It was pointed out in Section 7.3.2 that the method used to test the equality of group dispersions (Box, 1949) assumes the multivariate normality of the data. Although the descriptors selected were chosen to minimize such problems, reference to Section 3.7.1 indicates that there remain some departures from normality in the data for individual topodeme samples for some of the descriptors (Table 3.4).

Layard (1974) has demonstrated by Monte Carlo simulation methods that with non-normal data the

α-probabilties of the critical values of Box's M statistic are actually considerably larger than their nominal value. Hence as non-normality increases, so too does the likelihood of unjustifiably rejecting $H_0: \underline{\Delta}_k = \underline{\Delta}$, $k = 1...g$.

Campbell (1981) has proposed a number of graphical methods for comparing the covariance matrices of individual groups. Campbell treats the elements of group covariance matrices as the elements in one column of a q x g table, for q = 1...p variances or q = 1...p(p - 1)/2 correlations, and g = 1...m groups. Use is made of the variances and correlations since these statistics summarize the descriptor relationships in the data, and in addition can be transformed so as to have distributions more closely aproximating normality (Campbell, 1981).

The first of the methods suggested that is used here consists of plotting column elements against the corresponding row means, hence an "individual-average" (I-A) plot. Each group may be characterized by the slope of such a plot, and compared with other groups both by formal statistical tests of the equality of slopes, and by plotting their slopes against the corresponding column (group) means ("mean-slope" or M-S plot). Finally, deviations of column elements from the corresponding row means may be plotted against the normal probability scores of the pooled deviations. This will be referred to here simply as a

residual plot.

In applying these methods, variances of the ranged flower and fruit descriptors examined in the preceding sections (STYL, TCAL, WFL, LCAL, WFR) were coded by multiplication by 1000 and log-transformed, while their covariances were converted to correlations and arctanh transformed. These transformations were used to obtain variates as nearly normally distributed as possible (Sokal & Rohlf, 1969; Campbell 1981). Campbell (1981) suggests two possible row averages for use in the I-A and residual plots. The first of these is the corresponding element of the pooled covariance matrix, since this is calculated with a large number of degrees of freedom. Alternatively a robust estimate of the row mean may be obtained by trimming the extreme values in the sample and calculating the robust mean from the remaining elements (Brown & Forsythe, 1974). The results reported here for the correlations are based on the first method. Those for the variances, and for both sets of residual plots, are based on the use of robust means. Regressions and calculation of normal probability scores were done using the MINITAB package (Ryan, Joiner, & Ryan, 1976, 1981).

7.6.1 I-A and M-S Plots

Only three of the topodeme samples (T2, T3, T4) have significantly non-zero slopes for I-A plots of log variances ($p < 0.05$; Table 7.5). There are no significant differences among any of these slopes. Four topodeme samples (T2, T3, T4, T7) have significantly non-zero slopes for I-A plots of their arctanh correlations ($p < 0.05$; Table 7.6). All but two of the significant differences among the slopes indicated in Table 7.6 are due to comparisons with the smallest sample, T5 ($N = 7$), for which no trend is apparent in the relationship between its correlations and those of the pooled sample. Leaving aside comparisons with T5, the patterns of descriptor correlations in the remaining topodeme samples thus appear to be generally similar. A further indication of topodeme sample similarity with respect to descriptor correlations is obtained on examination of the residuals for the I-A plot regressions in Table 7.6. In comparisons between all pairs of topodeme samples (excluding T5) these residuals are significantly correlated ($p < 0.05$):

T2	0.863				
T3	0.687	0.677			
T4	0.820	0.640	0.760		
T6	0.921	0.773	0.657	0.795	
T7	0.640	0.639	0.560	0.647	0.805
	T1	T2	T3	T4	T6

Thus, given similar regression lines, it appears that the deviations from these lines also conform to a common pattern across not only both of the common crus-galli morphotypes (T1, T4, T2, T3, T6) but also the sample of C. punctata (T7) as well. That said, however, it can also be added that here the lowest correlations are predominantly between the residuals of T3 and of T7 and those of the other samples. Conversely, T1 and T6 tend to have the highest correlations with the other samples.

Differences among the topodéme samples are evident in the magnitude of both variances and correlations, as shown in the respective M-S plots (Fig. 7.1, 7.2). In Figure 7.1 the contrast is between the high mean log variance of the punctata sample (T7) and the appreciably smaller mean log variances of the crus-galli samples. In the average arctanh correlations for each topodeme sample the contrast seen is between T6 and all the other samples.

7.6.2 Residual Plots

The degree to which the pooled robust residuals for log variances and arctanh correlations for T1 - T7 represent a sample from a single, normally distributed population was assessed by the method of Ryan, Joiner and Ryan (1976; Ryan & Joiner, undated) described in Section 3.7.1. Critical values for the correlation between the robust residuals and

their normal probability scores for each sample size (q = 42 variances; q = 105 correlations; $\alpha = 0.05$) are 0.9726 and 0.9880, respectively.

The strongly linear trend (r = 0.980) of the residual plots for log variances (Fig. 7.4; Table 7.5) shows that the robust residuals (topodeme sample log variance minus corresponding robust mean) for all topodeme samples appear to represent a single population. None of the samples have data points with large departures from the regression line shown except T7 (C. punctata), for TCAL (Fig. 7.4f). The differences which are most conspicuous here are those in the magnitudes of the log variances (Fig. 7.2).

The trend of the pooled residuals for the arctanh correlations is noticeably less linear (r = 0.942) due mainly to the correlations for the smallest sample, T5 (Fig. 7.5d; Table 7.6). Repeating the analysis without T5 (not shown) results in a much more strongly linear relationship between the robust residuals and their normal scores (r = 0.989). Thus, leaving aside the sample in which correlations have been determined with only five degrees of freedom (T5), these results are strong evidence that the remaining topodeme samples are not markedly differentiated from one another with respect to the pattern of their descriptor correlations. To the extent that any differentiation is present, here it would appear to separate

T6 from the remainder of the sample, because of both the greater magnitude of its arctanh correlations (Fig. 7.3, 7.5e) and the departures from the linear trend in Fig. 7.5 of its robust residuals for the correlations between TCAL and STYL and between LCAL and TCAL (Table 7.6).

7.7 Discussion

Differences among topodeme sample covariance structures were investigated in part as a necessary prelude to statistical tests of the equality of topodeme sample mean vectors (Chapter 8). In addition, however, differences among groups with respect to patterns of descriptor variation and covariation are of intrinsic interest in view of the wide taxonomic range of the random sample in this study. It was expected that there might be differences in the relationships among the descriptors studied here.

With respect to the data for the leaf descriptors available for the 58 OTU subsample, the topodeme sample dispersions do not appear to differ significantly (Section 7.4; Table 7.1). The leaf descriptor variances and total variances ($\bar{y}$), and their covariance matrix determinants were found to be much higher and more uniform among T1 - T7 than those of the flower and fruit descriptors (Table 7.3a, 7.4a; see also Table 8.3).

The greater differences among the topodeme samples with respect to their dispersions for the flower and fruit descriptors appear to be a function of the variances more than of the patterns of descriptor covariation. The punctata topodeme sample (T7) is appreciably more variable than any of the crus-galli ones (T1 - T6), judging by the magnitudes of both the values of $\bar{y}$ (Table 7.3b) and the dispersion matrix determinants (Table 7.2a, c; Table 7.4). Rejection of $H_0: \underset{\sim}{\Delta}_k = \underset{\sim}{\Delta}$, k = 1...g for the six crus-galli topodeme samples (Table 7.2b) is probably due to the very small determinant obtained for T5. This in turn resulted from the extremely low variances for TCAL, WFL and WFR in this small sample (Fig. 7.4d; Van Valen, 1978). Likewise, the relatively low variances and also the high correlations of the descriptors in sample T6 (Fig. 7.4e, 7.5e) probably account for the small determinant of its dispersion matrix (Table 7.4). A posteriori comparison of the crus-galli topodeme samples with respect to $\bar{y}$, however, is not affected by these factors, and it is clear that while $\bar{y}$ is significantly higher for T7, samples T1 - T6 are not distinguishable from one another with respect to this parameter (Table 7.3b).

Some evidence was also obtained of differentiation among the topodeme samples with respect to the pattern of descriptor covariation. Although log variance I-A plot

slopes did not differ significantly among the topodeme samples (Table 7.5), the corresponding M-S plot indicates some differentiation of T2 and T3 from the other samples (Fig. 7.2), corresponding to a tendency for descriptor variances for these samples to be close to the robust averages for the entire random sample (Table 7.5). The other topodeme samples either tend to have smaller variances (T4) or else no significant trehd is present (T1, T5-T7).

Differences in the pattern of descriptor covariation as revealed by significantly different slopes in the I-A plots of arctanh correlations distinguish T3 and T7 from the other samples (Fig. 7.3). This corresponds to the way in which their correlations tend to be consistently closer to those of the entire 93 OTU crus-galli random sample than are the ones of the other topodeme samples (Table 7.6).

The most conspicuously differentiated topodeme sample, however, is T6 by reason of its very high descriptor correlations (Fig. 7.3, 7.5; Table 7.6). This is a reflection particularly of the tendency of OTUs in this topodeme to consistently have few styles and little or no toothing on their calyx lobes (Table 3.11; Table 7.6).

The small degree of differentiation with respect to their covariance structure of the seven topodeme samples examined here is a little surprising in view of the

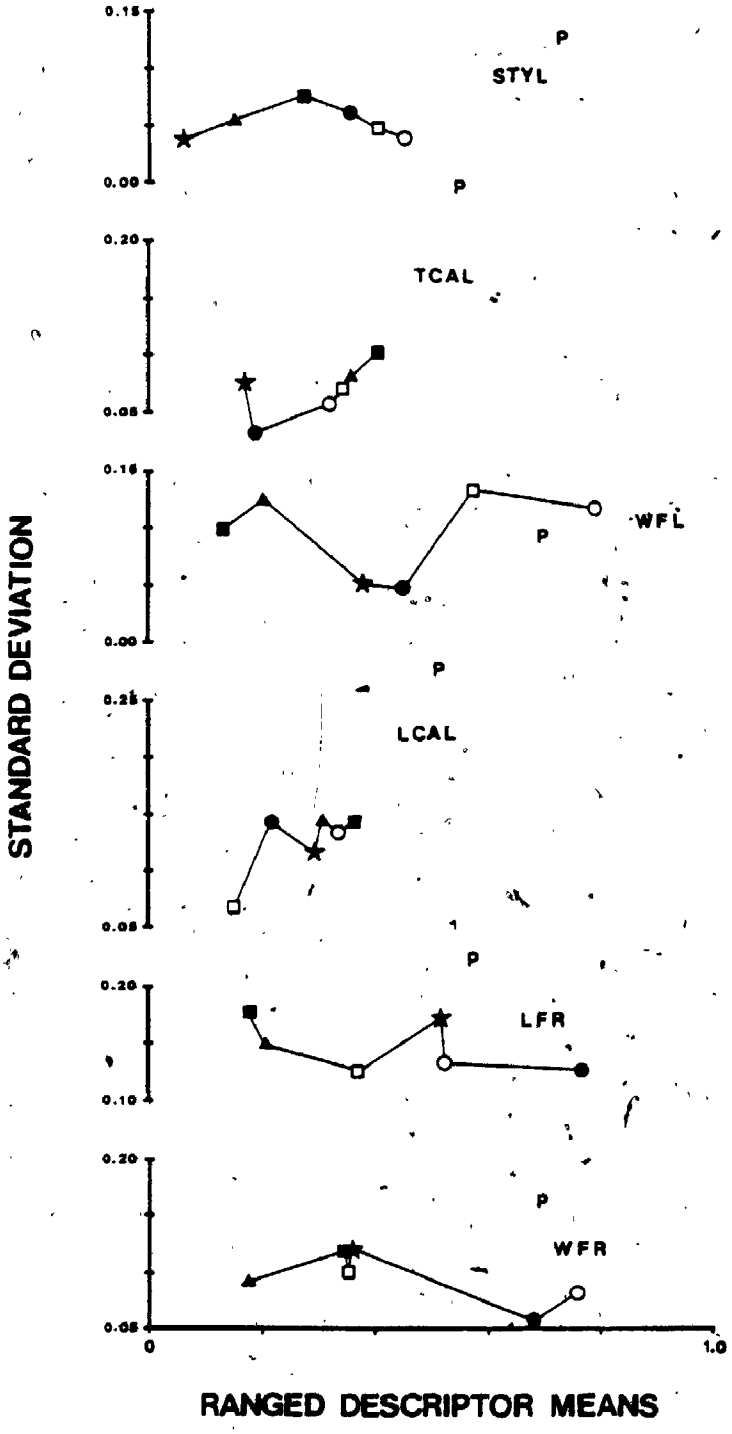
taxonomic diversity of the sample. However, repetition of analyses such as these with a suite of descriptors which samples more of the phenotype might well lead to a picture of more complex descriptor covariation. The six descriptors that have been studied here are in fact largely measures of size and number for attributes of flowers and fruit whose relationships (e.g. shape), while they do distinguish both the topodeme samples and the taxonomic groups in this study (Chapter 8) do not do so as dramatically as do the multistate descriptors examined earlier (Section 6.3).

Beside their importance to formal tests of the differentiation of topodeme samples from one another, comparisons of the degree of topodeme sample variabiilty are of interest in their own right (Day & Fisher, 1937). One area of concern in evolutionary biology at present is to achieve an understanding of the significance of sexual reproduction in all its diverse expressions (Williams, 1975; Smith, 1978; Charlesworth, 1980; Lloyd, 1980). This may be seen as an understanding of the different balances which may be realized between maximizing the faithful transmission of an individual's genotype to offspring and maximizing the reproductive success of those offspring by enabling acquisition of advantageous genes or gene combinations (from other individuals (Bremermann, 1980 and undated; Lloyd, 1980).

The comparisons made here document only phenotypic variation and variability. Assessment of the purely genetic component of these parameters requires more detailed investigation either by means of common-garden experiments (Barrett, 1982; Warwick & McNeill, 1982), or else by more direct sampling of the genome. The latter approach is exemplified by methods assaying protein polymorphism (Gottlieb, 1977) and variation in nucleotide sequence and copy number of particular genes (Walbot, Beachy & Yao, 1980; Schaal, 1982).

Nevertheless, integration of the results of the analyses described in this chapter and in Chapters 5, 6 and 8 with the results of studies of the reproductive behavior of individuals in each of the topodeme samples (Chapter 10) provides a basis for constructing hypotheses about the structure and history of Crataegus topodemes (Chapter 11).

Figure 7.1 Plot of descriptor standard deviations against the corresponding means, for ranged data of topodeme samples T1 - T7. The line connecting data points for T1 - T6 shows the trend in standard deviations with increasing mean. See Table 3.1 for explanation of descriptor abbreviations. The 10-stamen morphotype of C. crus-galli (T2, T3, T5, T6) is represented by solid symbols, the 20-stamen one (T1, T4) by open symbols and C. punctata (Site 5; T7) by the letter P. Individual sites are coded as follows: Site 1, squares; Site 2, circles; Site 3 (T3), triangles; and Site 4 (T6), stars.


STANDARD DEVIATION
0.15
0.00
P
STYL
P
0.20
0.05
TCAL
0.15
0.00
P
WFL
P
0.25
0.05
LCAL
P
0.20
0.10
LFR
0.20
0.05
P
WFR
0
1.0
RANGED DESCRIPTOR MEANS

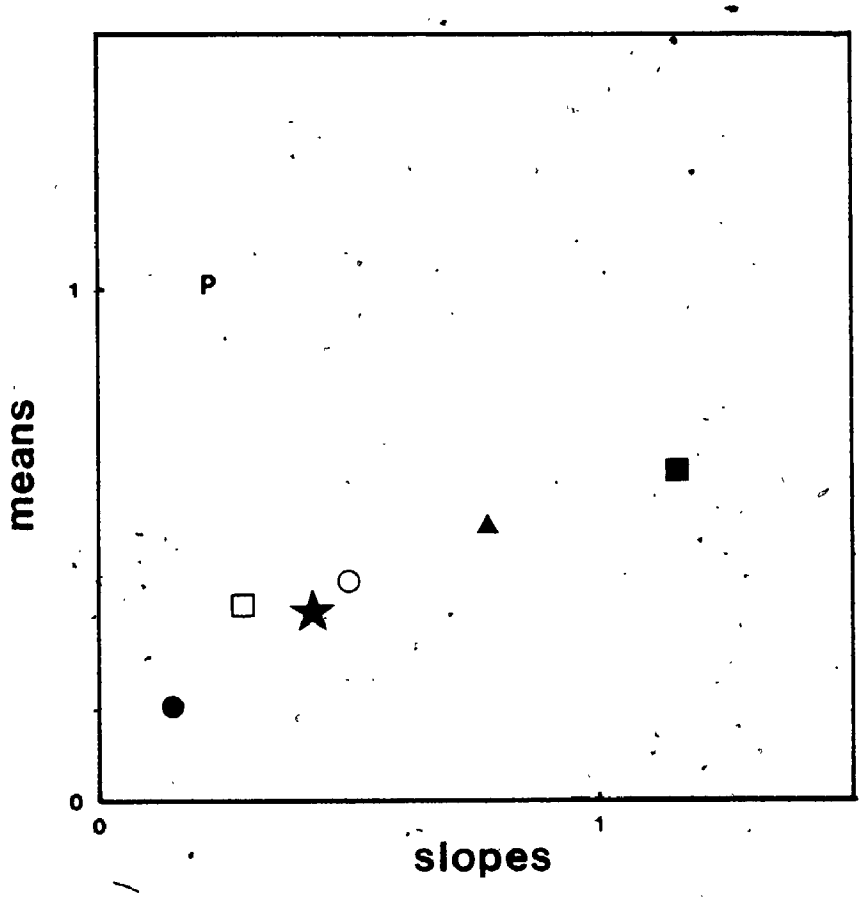


Figure 7.2 Mean-slope plot of log transformed variances of six ranged descriptors (STYL, TCAL, WFL, LCAL, LFR, WFR; Table 3.1) for seven topodeme samples (T1 - T7) making up the 111 OTU random sample. Means are the mean log variance for each topodeme sample. Slopes are those of the corresponding individual-average plot (see Table 7.4). See text for details. See Figure 7.1 for explanation of symbols.

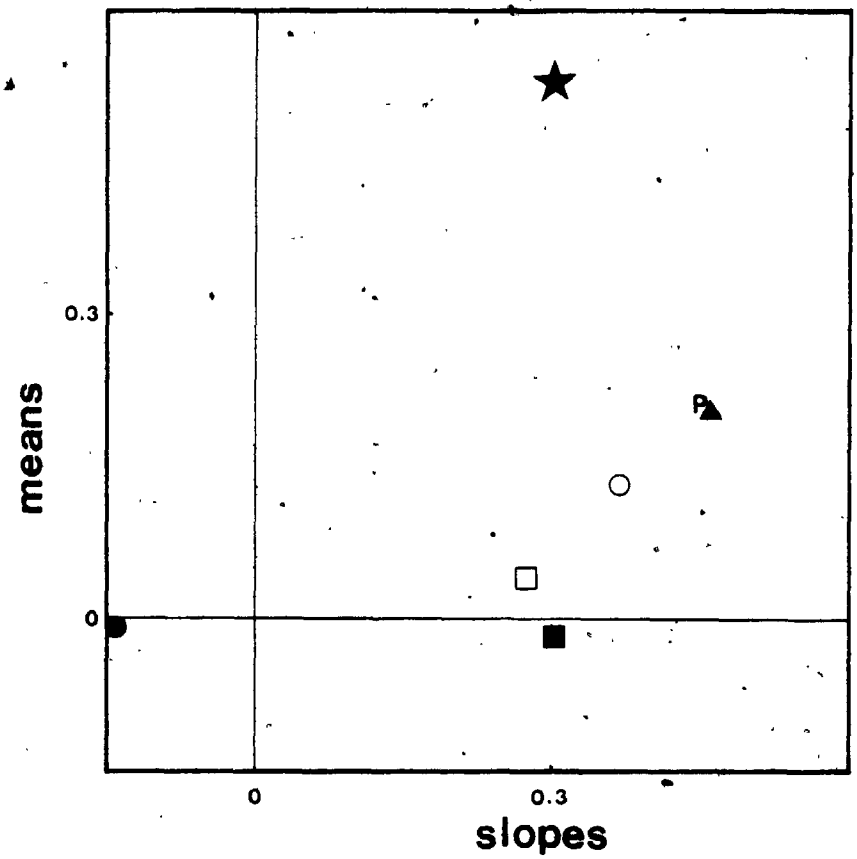
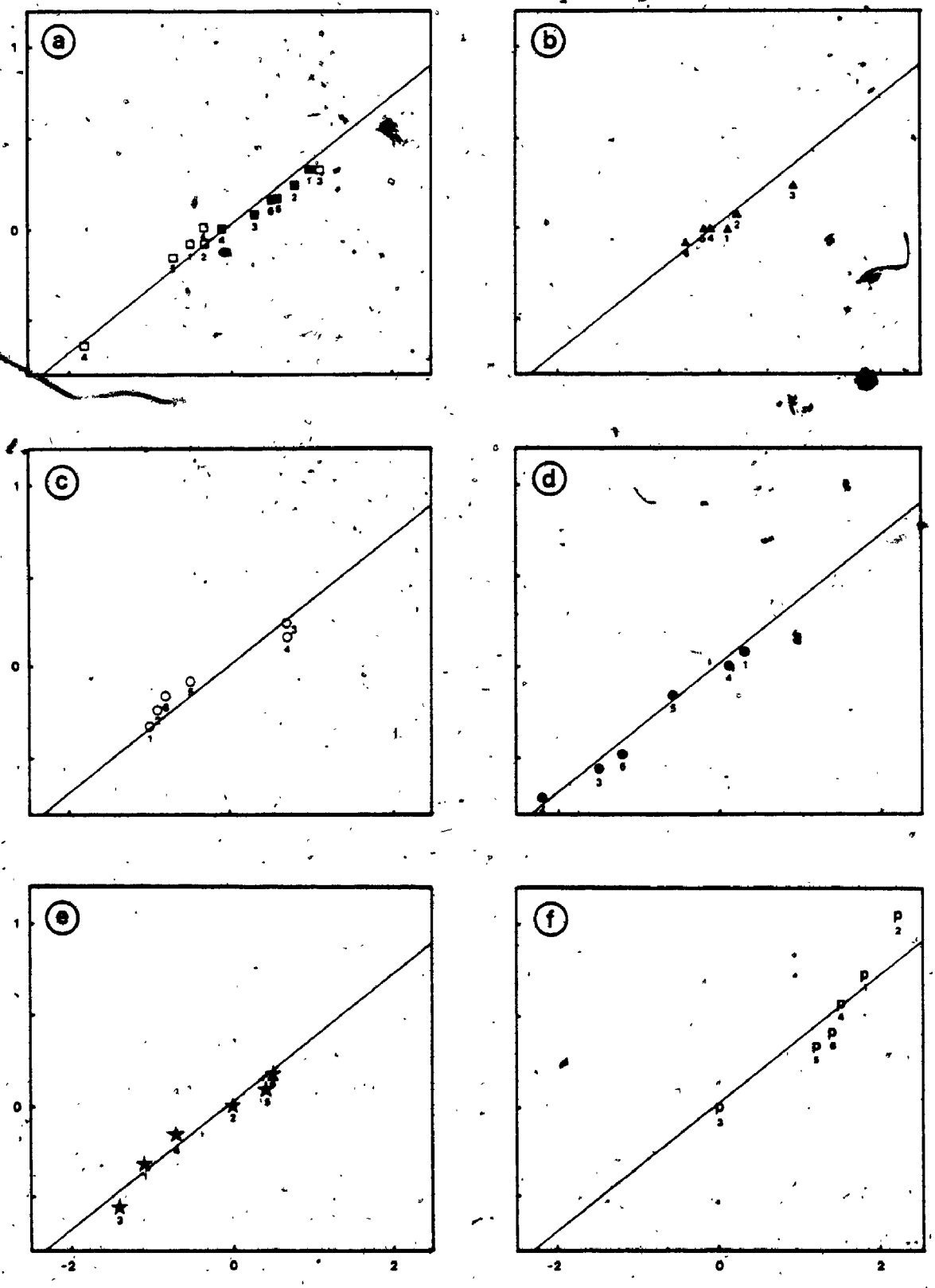


Figure 7.3 Mean-slope plot of the arctanh transformed correlations between six ranged descriptors (STYL, TCAL, WFL, LCAL, LFR, WFR; Table 3.1) for seven topodeme samples (T1 - T7) making up the 111 OTU random sample. Means are the mean arctanh correlation for each topodeme sample. Slopes are those of the corresponding individual-average plot (see Table 7.5). See text for details. See Figure 7.1 for explanation of symbols.

Figure 7.4 Residual plots for log variances for seven topodeme samples (T1 - T7). Ordinate: robust residuals. Abcissa: normal probability scores of the robust residuals for all seven samples. The line shown is the corresponding regression line. See text and Table 7.5 for details. Topodeme samples are plotted separately as follows: (a) T1 (open symbols) and T2 (solid symbols); (b) T3; (c) T4; (d) T5; (e) T6; (f) T7. Descriptors are represented as follows: (1) STYL; (2) TCAL; (3) WFL; (4) LCAL; (5) LFR; (6) WFR. Scales for each axis are the same in a - f.


a
b
c
d
e
f
1
0
-2
0
2
p1
p2
p3
p4
p5
p6

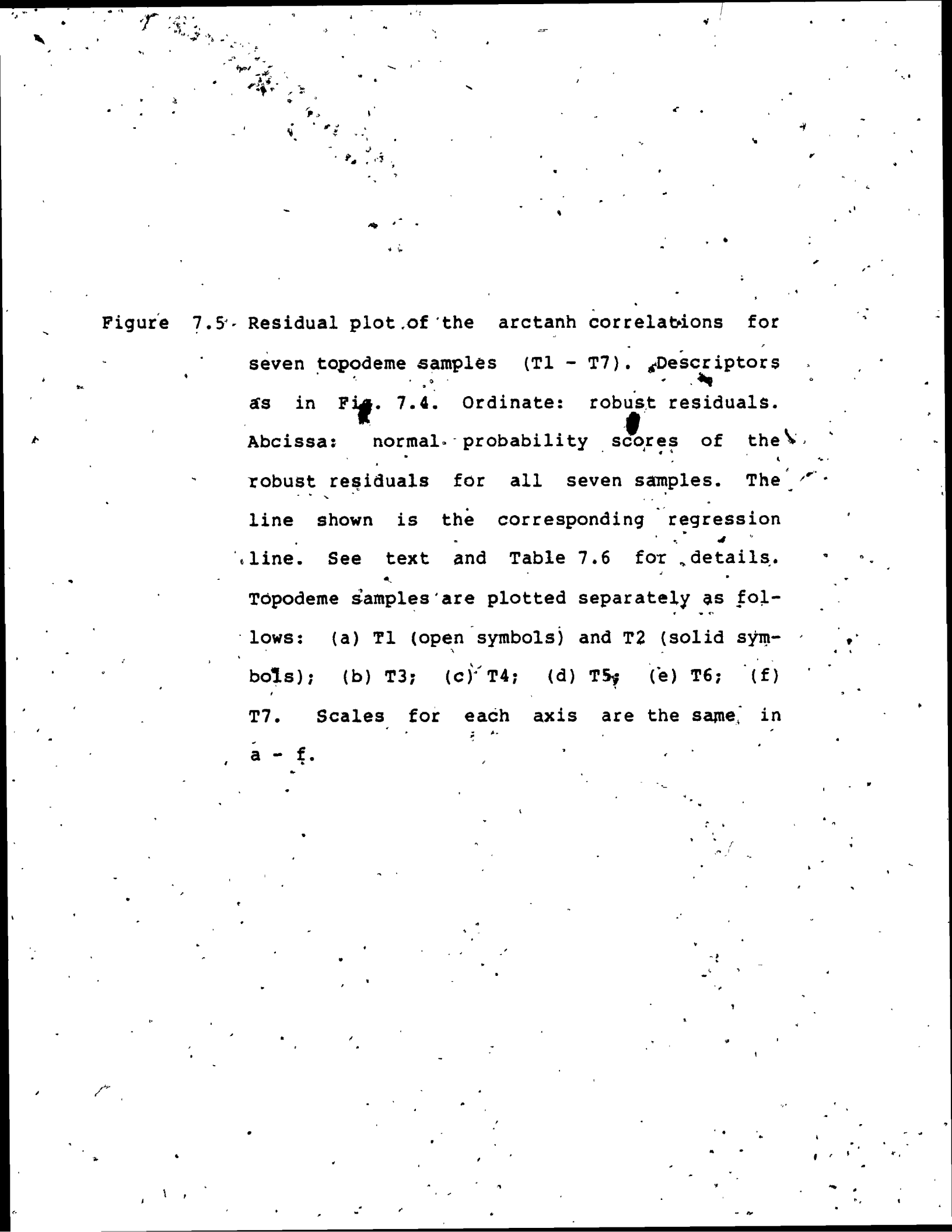


Figure 7.5 Residual plot of the arctanh correlations for seven topodeme samples (T1 - T7). Descriptors as in Fig. 7.4. Ordinate: robust residuals. Abcissa: normal probability scores of the robust residuals for all seven samples. The line shown is the corresponding regression line. See text and Table 7.6 for details. Topodeme samples are plotted separately as follows: (a) T1 (open symbols) and T2 (solid symbols); (b) T3; (c) T4; (d) T5; (e) T6; (f) T7. Scales for each axis are the same in a - f.

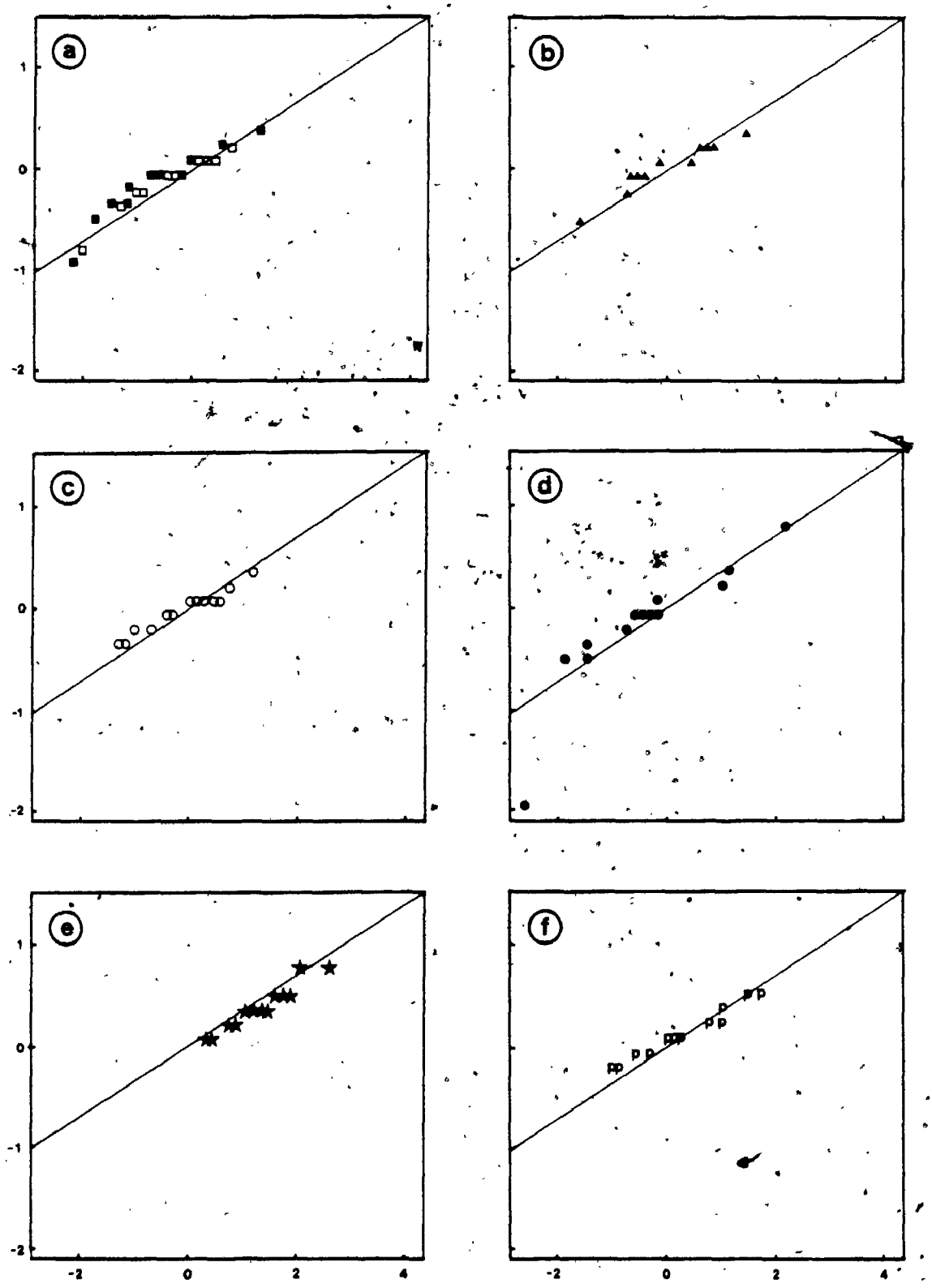

a
b
c
d
e
f
1
0
-1
-2
-2
0
2
4

Table 7:1 Results of testing $H_0: \underset{\sim}{\Delta}_k = \underset{\sim}{\Delta}$, $k = 1...g$, using ranged leaf descriptor data (1) and different numbers of groups, defined by the criteria indicated. The test statistic is the F value equivalent to the Bartlett-Box M statistic (Cooley & Lohnes, 1971). See text for details.

	M	n_1	n_2	F_s
(a) 2 Groups				
(Stamen Number)	89.065	21	10507	3.744***
(b) 5 Groups				
(Sites)	224.880	84	2755	1.857***
(c) 7 Groups				
(Topodemes)	240.440	126	3319	1.192ns

(1) Descriptors X, Y, Z, NUMSEC, ANGSEC and TEETH, scored for the 58 OTU subsample.

*** $p < 0.001$

ns $p > 0.05$

Table 7.2 Results of testing $H_0: \underline{\Delta}_k = \underline{\Delta}$, $k = 1...g$, using ranged flower and fruit data, different numbers of OTUs, and different subsets of descriptors. The groups are the topodeme samples. The test statistic is the F value equivalent to the Bartlett-Box M statistic (Cooley & Lohnes, 1971). See text for details.

	M	n_1	n_2	F_s
(a) 58 OTU Subsample, 7 Groups				
6 Descriptors (1)	276.766	126	3319	1.373**
(b) 93 OTU Subsample				
6 Descriptors (1)	198.292	105	3937	1.442**
5 Descriptors (2)	137.283	75	4084	1.475**
4 Descriptors (3)	92.305	50	4370	1.563**
(c) 111 OTU Sample				
6 Descriptors (1)	343.385	126	5185	2.125***
5 Descriptors (2)	227.452	90	5370	2.069***
4 Descriptors (3)	144.312	60	5730	2.061***

Table 7.2 Cont.

(1) Descriptors STYL, TCAL, WFL, LCAL, LFR, WFR.

(2) Descriptors STYL, WFL, LCAL, LFR, WFR.

(3) Descriptors WFL, LCAL, LFR, WFR.

** $0.01 > p > 0.001$

*** $p < 0.001$

Table 7.3 Comparison of individual topodeme sample total variances (transformed to $\bar{y}$ and averaged for each sample as described in text) for ranged leaf, and flower and fruit descriptors.

(a) Leaf Descriptors (1), 58 OTU Subsample

Sample	T1	T2	T3	T4	T5	T6	T7
N	8	8	8	10	7	9	8
$\bar{y}$ (2)	0.28	0.23	0.32	0.30	0.23	0.29	0.35

(b) Flower and Fruit Descriptors (3), 111 OTU Sample

Sample	T1	T2	T3	T4	T5	T6	T7
N	10	17	16	24	7	19	18
$\bar{y}$ (4)	0.21a	0.25a	0.26a	0.21a	0.19a	0.24a	0.44

(1) Descriptors X, Y, Z, NUMSEC, ANGSEC, TEETH.

(2) Values of $\bar{y}$ not found to be significantly different by one way analysis of variance ($F_s[6,51] = 1.02$; $p > 0.05$).

(3) Descriptors STYL, TCAL, WFL, LCAL, LFR, WFR.

(4) Values of $\bar{y}$ found to be significantly different by one way analysis of variance ($F_s[6,104] = 6.99$). Values of $\bar{y}$ followed by the same letter are not significantly different ($p < 0.05$) according to the Student-Newman-Keuls test (Sokal & Rohlf, 1969).

Table 7.4 Comparison of the determinants of individual topodeme sample covariance matrices, for ranged leaf, and flower and fruit descriptors.

(a) Leaf Descriptors (1), 58 OTU Subsample

Sample	T1	T2	T3	T4	T5	T6	T7
N	8	8	8	10	7	9	8
Det(2)	16.7	2.2	12.3	20.0	1.5	0.5	8.9

(b) 6 Flower and Fruit Descriptors (3), 58 OTU Subsample

Sample	T1	T2	T3	T4	T5	T6	T7
N	8	8	8	10	7	9	8
Det(2)	0.0015	0.0001	0.0003	0.0026	0.00001	0.0001	0.0529

Table 7.4 Cont.

(c) 6 Flower and Fruit Descriptors (3), 111 OTU Sample

Sample	T1	T2	T3	T4	T5	T6	T7
N	10	17	16	24	7	19	18
Det(2)	0.029	0.817	0.169	0.057	0.00006	0.003	95.03

5 Flower and Fruit Descriptors (4)

Sample	T1	T2	T3	T4	T5	T6	T7
N	10	17	16	24	7	19	18
Det(2)	1.66	29.96	10.81	3.17	0.035	0.54	799.3

4 Flower and Fruit Descriptors (5)

Sample	T1	T2	T3	T4	T5	T6	T7
N	10	17	16	24	7	19	18
Det(6)	0.039	0.148	0.149	0.060	0.001	0.019	2.860

(1) Descriptors X, Y, Z, NUMSEC, ANGSEC, TEETH.

(2) Determinant x 10^{15}.

(3) Descriptors STYL, TCAL, WFL, LCAL, LFR, WFR.

(4) Descriptors STYL, WFL, LCAL, LFR, WFR.

(5) Descriptors WFL, LCAL, LFR, WFR.

(6) Determinant x 10^{10}.

Table 7.5 Comparison of the log transformed variances of the individual topodeme random samples with each other and with the robust mean log variances, for the descriptors indicated (see Table 3.1 for explanation of abbreviations). Slopes shown are those for the regression of the robust means on the corresponding topodeme sample variances (I - A plot). Significantly non-zero slopes are underlined ($\alpha = 0.05$). None of the slopes differ significantly from any other. See text for details. Note that the variances were calculated from the ranged OTU descriptor scores scores, multiplied by 1000 prior to transformation.

Sample	T1	T2	T3	T4	T5	T6	T7	(1)
STYL	-0.049	0.362	0.077	-0.214	0.156	-0.244	0.796	0.067
TCAL	0.487	0.817	0.608	0.325	-0.186	0.559	1.576	0.559
WFL	0.854	0.594	0.794	0.739	-0.060	0.013	0.541	0.536
LCAL	0.006	0.646	0.646	0.830	0.653	0.462	1.220	0.648
LFR	0.466	0.765	0.615	0.507	0.474	0.736	0.971	0.620
WFR	0.554	0.690	0.474	0.371	0.055	0.703	0.967	0.558
Slope	0.293	<u>1.150</u>	<u>0.771</u>	<u>0.505</u>	0.150	0.430	0.210	

(1) Robust mean log variance calculated from the five median values for each descriptor.

Table 7.6 Comparison of the correlation structure of the dispersion matrices of the individual topodeme random samples with each other and that of the 93 OTU dispersion matrix. Entries are arctanh transformed correlations between the descriptors shown at the left (1). Significant correlations ($\alpha = 0.05$) are underlined. The slopes indicated are those for the regression of the 93 OTU correlations on the corresponding topodeme correlations (I - A plot). Significantly non-zero slopes are underlined ($\alpha = 0.05$). See text for details.

Sample	T1	T2	T3	T4	T5	T6	T7	
N	10	17	16	24	7	19	18	93
TCAL-STYL	0.157	-0.254	-0.345	0.130	0.017	<u>0.958</u>	<u>0.545</u>	0.115
WFL-STYL	-0.585	-0.223	0.379	0.362	0.233	0.471	<u>0.732</u>	<u>0.310</u>
WFL-TCAL	-0.094	0.508	0.264	-0.218	0.086	0.327	0.156	-0.026

Table 7.6 Cont.

Sample	T1	T2	T3	T4	T5	T6	T7	
LCAL-STYL	0.154	-0.099	0.354	0.323	0.199	0.788	0.173	0.005
LCAL-TCAL	0.239	0.112	0.134	-0.193	-0.317	0.915	0.400	0.048
LCAL-WFL	0.443	0.136	0.289	0.792	1.242	0.477	0.163	0.027
LFR-STYL	0.181	0.153	-0.124	-0.066	-0.116	0.325	-0.202	0.108
LFR-TCAL	-0.171	-0.207	0.059	-0.080	-0.472	0.306	-0.070	-0.113
LFR-WFL	-0.087	0.102	0.490	-0.007	-0.480	0.202	0.307	0.244
LFR-LCAL	-0.342	-0.980	0.117	0.015	-0.188	0.263	-0.073	-0.060
WFR-STYL	0.230	0.220	0.111	0.036	0.494	0.428	0.109	0.320
WFR-TCAL	-0.070	-0.141	-0.157	-0.268	0.374	0.503	0.070	-0.028
WFR-WFL	-0.166	0.154	0.370	0.198	0.030	0.255	0.528	0.355
WFR-LCAL	-0.217	-0.573	0.154	0.102	-0.154	0.285	-0.364	0.013
WFR-LFR	0.937	0.877	1.054	0.928	-1.043	1.423	0.660	0.710

Table 7.6 Cont.

Sample	T1	T2	T3	T4	T5	T6	T7
Slope(2)	0.273	0.301	0.456	0.368	-0.146	0.300	0.445
	ab	cde	acf	g	b,d, f-i	h	e,i

(1) See Table 3.1 for explanation of abbreviations.

(2) Samples with the same letter have significantly different slopes ($\alpha = 0.05$).

CHAPTER EIGHT

EVALUATION OF GROUP STRUCTURE

8.1 Introduction

Both the total sample of 160 OTUs (Chapter 5) and the subsample of 60 OTUs (Section 6.3) have been subjected to cluster analyses employing different sorting algorithms, resemblance functions, and (with the smaller sample) different data sets. However, by only comparing different cluster analyses, the possibility remains that clustering itself is inappropriate, and that the methods used merely subdivide a single homogeneous cluster of OTUs. One check on this is the ordination of the sample, so as to determine visually whether or not appreciable discontinuities are present in the space of the first few axes (Cormack, 1971). This has been done already (Section 6.2), and the results suggest that there are indeed discontinuities present in both the 160 and 60 OTU samples. The degree of discontinuity varies, however, with both data transformation (Section 6.2.1) and data set (Section 6.2.2), while the groups distinguished appeared to vary somewhat with data set as well (Section 6.2.2; Section 7.4).

The same discontinuities in the 160 and 60 OTU samples were detected by the corresponding cluster analyses, although the degree of discontinuity between clusters varied with the resemblance function used (Chapter 5; Section 6.3). Also, the hierarchical relationships among the clusters changed, notably with sorting algorithm (patterns A - C; Chapter 5; Section 6.3). Nevertheless, with both samples, and with most combinations of data type and method (for the smaller sample), the cluster analyses suggest that the OTUs may readily be grouped according to site and taxon, that is, by topodeme sample in the case of the randomly sampled OTUs.

The present chapter is concerned with testing the validity of these groupings. This is done first in the context of formal statistical tests based on the multivariate analysis of variance introduced in Chapter 7. Next, canonical analysis is presented as a general technique of which the multivariate analysis of variance and related techniques represent applications. Canonical analyses are used to provide further information about the interrelationships between the data sets available and about the way in which they discriminate the topodeme samples. This is followed by two sections, the first of which examines the performance of certain of the canonical analyses, while the second presents the results of an

alternative method of examining the distinctness of groups. A last section is devoted to an example of regression techniques that provides additional evidence of topodeme sample differentiation.

There are a number of additional methods of testing the group structure of data, based on analysis of contingency tables (Strahler, 1978; Feoli & Orlóci, 1979; Gittins, 1979; Gokhale, 1979). Since the data set in use here is made up predominantly of continuous descriptors it was felt that methods such as these were unnecessary.

Other a posteriori approaches to the evaluation of group structures exist but have not been employed here. Both parametric (Sneath, 1977, 1979a,b; Hill, 1980) and non-parametric (Sneath, 1979c) tests have been proposed for evaluating the extent to which two groups of OTUs implied by a phenogram may be distinguished from one another, on the basis of between-OTU distances or descriptor-state frequency distributions. Conversely, Beshir (1975; Orlóci & Beshir, 1976) proposed a test of group homogeneity (i.e. of the absence of non-random variation within a group) based on the comparison of observed and expected distributions of between-OTU distances.

Since the partitions of interest here are not necessarily binary, nor are they ones implied by a

particular phenogram, such methods are less useful than the ones discussed below. Likewise the comparision of n x n matrices of phenogram descriptors (Section 6.3) with the original resemblance matrix (Rohlf & Sokal, 1981) represents not so much an evaluation of group structure itself as it does an evaluation of the fidelity with which a resemblance structure has been reproduced by a given phenogram. In view of the consistency with which topodeme samples appeared as groups in the cluster analyses, such comparisons seem unnecessary.

8.2 Multivariate Analysis of Variance

Tables 8.1 and 8.2 present the results of testing the null hypothesis of the equality of group mean vectors, $H_0: \underset{\sim}{\mu}_k = \underset{\sim}{\mu}$, k = 1...g, that was described in Section 7.3. The test statistic used is Wilks' Λ (Wilks, 1932),

$$\Lambda = |\underset{\sim}{W}|/|\underset{\sim}{T}|$$

and the corresponding F approximation given by Rao (1952; Cooley & Lohnes, 1971; Green, 1978). Although the null hypothesis is rejected in every case, only with the leaf data (Table 8.1) and a hypothesis of seven groups (topodeme samples) can formal statistical significance be attached to the inequality of group mean vectors. Only for this data set and value of g is the prior null hypothesis of the equality of group dispersions, $H_0: \underset{\sim}{\Delta}_k = \underset{\sim}{\Delta}$, k = 1...g accepted

(Table 7.1). With the other data sets and partitions of the sample the groups have already been shown to differ with respect to their dispersion matrices (Table 7.1, 7.2), so that comparison of their centroids as described here is unjustified.

Table 8.3 gives univariate F ratios for the twelve descriptors in Table 8.1 and 8.2. These provide an indication of the descriptors contributing most to group discrimination (Cooley & Lohnes, 1971).

Table 8.1 Results of testing $H_0: \underset{\sim}{\mu}_k = \underset{\sim}{\mu}$, k = 1...g, using ranged leaf descriptor data (1) available for the 58 OTU subsample and different numbers of groups, defined by the criteria indicated (a-c). The test statistic is the F value equivalent to Wilks' Λ (Cooley & Lohnes, 1971). See text for details.

	Λ	n_1	n_2	F_s
(a) 2 Groups (2) (Stamen number)	0.1964	6	51	34.78***
(b) 5 Groups (2) (Sites)	0.0213	24	168	14.14***
(c) 7 Groups (3) (Topodemes)	0.0023	36	204	16.95***

*** $p < 0.001$

(1) Descriptors X, Y, Z, NUMSEC, ANGSEC and TEETH; see Table 3.3 for explanation of descriptor abbreviations.

(2) Partition of the sample for which $H_0: \underset{\sim}{\Delta}_k = \underset{\sim}{\Delta}$, k = 1...g rejected (Table 7.1).

(3) Partition of the sample for which $H_0: \underset{\sim}{\Delta}_k = \underset{\sim}{\Delta}$, k = 1...g accepted (Table 7.1).

Table 8.2 Results of testing $H_0: \underset{\sim}{\mu}_k = \underset{\sim}{\mu}$, k = 1...g, using ranged flower and fruit descriptor data, different numbers of OTUs (a-c) and different subsets of descriptors (1-3). The groups are topodeme samples. The test statistic is the F value equivalent to Wilks' Λ (Cooley & Lohnes, 1971). See text for details. Note that for all partitions of the sample shown below $H_0: \underset{\sim}{\Delta}_k = \underset{\sim}{\Delta}$, k = 1...g was rejected (Table 7.2).

	Λ	n_1	n_2	F_s
(a) 58 OTU Subsample, 7 Groups (T1 - T7)				
6 Descriptors (1)	0.0010	36	204	21.69***
(b) 93 OTU Subsample, 6 Groups (T1 - T6)				
6 Descriptors (1)	0.0025	30	330	38.13***
5 Descriptors (2)	0.0037	25	309	43.58***
4 Descriptors (3)	0.0132	20	279	37.62***
(c) 111 OTU Sample, 7 Groups (T1 - T7)				
6 Descriptors (1)	0.0035	36	437	31.92***
5 Descriptors (2)	0.0040	30	402	39.92***
4 Descriptors (3)	0.0201	24	353	30.41***

Table 8.2 Cont.

*** $p < 0.001$

(1) Descriptors STYL, TCAL, WFL, LCAL, LFR, WFR; see Table 3.1 for explanation of descriptor abbreviations.

(2) Descriptors STYL, WFL, LCAL, LFR, WFR.

(3) Descriptors WFL, LCAL, LFR, WFR.

Table 8.3 Univariate F-ratios for 12 raw descriptors for 7 topodeme random samples (T1 - T7) for which $H_0: \underset{\sim}{\mu}_k = \underset{\sim}{\mu}$, k = 1...7 has been rejected (Table 8.1, 8.2).

Descriptor(1)	Between MS	d.f.	Within MS	d.f.	F
STYL(2)	8.44	6	0.05	104	173.86***
TCAL(2)	1.45	6	0.08	104	17.65***
WFL(3)	4.03	6	0.04	104	108.44***
LCAL(2)	1.14	6	0.17	104	6.58***
LFR(3)	9.63	6	0.46	104	20.92***
WFR(3)	36.27	6	0.54	104	67.12***
X(4)	35.69	6	3.53	51	10.12**
Y(4)	26.40	6	3.95	51	6.69*
Z(4)	94.41	6	13.82	51	6.83*
NUMSEC(4)	5.97	6	4.68	51	18.32***
ANGSEC(4)	179.90	6	4.68	51	38.41***
TEETH(4)	14.88	6	0.37	51	39.72***

* $0.05 > p > 0.01$ ** $0.01 > p > 0.001$ *** $p < 0.001$

(1) See Table 3.1, 3.3 for explanation of abbreviations.

(2) Variances significantly heterogeneous ($p < 0.01$) according to approximate Fmax test for unequal samples.

(3) Variances not significantly heterogeneous ($p > 0.05$) according to approximate Fmax test for unequal samples.

(4) Variances not significantly heterogeneous ($p > 0.05$) according to Fmax test.

8.3 Canonical Analysis

Where more than one set of data is available for a sample of OTUs, canonical analysis is a powerful tool for exploring the relationships between data sets (Gittins, 1979). The following outline of canonical analysis methodology is based principally on the expositions of Gittins (1979) and Green (1978).

Canonical analysis is comparable to PCA, but with certain differences. Whereas in PCA the total variance associated with a sample described by a single set of descriptors is partitioned among successive orthogonal axes (eigenvectors of the sample dispersion matrix), in canonical analysis the covariance between two (or more) data sets is examined. In a sample of OTUs the jth OTU is represented by a partitioned vector $\underset{\sim}{z}_j$ consisting of subvectors $\underset{\sim}{x}_j$ and $\underset{\sim}{y}_j$, where x_{hj} and y_{ij} represent elements of the first and second data sets respectively ($h = 1...p$, $i = 1...q$). The correlation structure of the sample is described by

$$\underset{\sim}{R} = 1/n\sum_j \underset{\sim}{z}_j \underset{\sim}{z}_j', \quad j = 1...n$$

$$= \begin{array}{|c|c|} \hline \underset{\sim}{R}_{xx} & \underset{\sim}{R}_{xy} \\ \hline \underset{\sim}{R}_{yx} & \underset{\sim}{R}_{yy} \\ \hline \end{array}$$

Linear transformations are found for each set of descriptors

which maximize the correlation between the transformed descriptors. This is done by solving the eigenstructure of the nonsymmetric matrix product

$$\underset{\sim}{F} = \underset{\sim xx}{R^{-1}}\underset{\sim xy}{R}\ \underset{\sim yy}{R^{-1}}\underset{\sim yx}{R}$$

This results in the diagonal matrix $\underset{\sim}{D}$ of s eigenvalues (λ_k, k = 1...s, s = min(p,q)) and the matrix $\underset{\sim}{V}$ of the corresponding eigenvectors. The columns of $\underset{\sim}{V}$ may be used as a set of canonical weights for the first set of descriptors, and be tranformed to give the canonical weights for the second set as follows:

$$\underset{\sim}{A} = \underset{\sim yy}{R^{-1}}\underset{\sim yx}{R}\ \underset{\sim\sim}{V}D^{-\frac{1}{2}}$$ (program BMD-P6M; Dixon & Brown, 1979).

Alternatively (Cooley & Lohnes, 1971; Green, 1978), the columns of $\underset{\sim}{V}$ may be scaled so as to give matrices of canonical weights $\underset{\sim}{B}$ (for the first set) and $\underset{\sim}{A}$ (for the second set) that meet the requirements

$$\underset{\sim}{B}'\underset{\sim yy}{R}\underset{\sim}{B} = \underset{\sim}{I}$$

$$\text{and } \underset{\sim}{A}'\underset{\sim xx}{R}\underset{\sim}{A} = \underset{\sim}{I}.$$

These canonical weights are used to calculate the transformation of the original data to scores on s pairs of axes or canonical variates, one member of each pair corresponding to the first data set, and the other to the second data set. The rescaling of the eigenvectors described above has the effect of producing canonical

variates the scores on which have unit variances.

The following two sections (8.4, 8.5) discuss the two principal applications of canonical analysis: canonical correlation analysis (CCA; Section 8.4) and canonical variates analysis (CVA; Section 8.5). As above, the description of each of these methodologies is based on the treatments by Gittins (1979), Green (1978), and Cooley and Lohnes (1971).

8.4 Canonical Correlations Analysis

The scores for individual OTUs on each pair of canonical variates are calculated from the scores for each set of descriptors, with or without standardization, as

$$\underset{\sim}{X}\underset{\sim}{V} = \underset{\sim}{U} \quad (\text{or } \underset{\sim}{X}\underset{\sim}{B} = \underset{\sim}{U})$$

$$\text{and } \underset{\sim}{Y}\underset{\sim}{A} = \underset{\sim}{T}.$$

Several statistics are of interest. The diagonal matrix of canonical correlations $\underset{\sim}{M}$ describes the correlations between members of successive pairs of canonical variates ($\underset{\sim}{U}_k$, $\underset{\sim}{T}_k$ for k = 1...s), where

$$\underset{\sim}{M} = \underset{\sim}{V}'\underset{\sim}{R}_{xy}\underset{\sim}{A} = \underset{\sim}{D}^{\frac{1}{2}}$$

The kth canonical correlation ($m_k = \lambda_k^{\frac{1}{2}}$) is interpretable as the proportion of variance in $\underset{\sim}{U}_k$ predictable from or common to $\underset{\sim}{T}_k$, and vice-versa (Gittins, 1979).

The significance of canonical correlations can also be tested by means of Wilks' Λ statistic (Section 8.2) and its χ^2 approximation (Bartlett, 1947). Wilks' Λ has already been defined as the ratio of the determinants of pooled within-groups and total sample dispersion matrices, in the context of the multivariate analysis of variance (Section 8.2). It can be shown (Green, 1978) that Λ can be expressed in terms of the squared canonical correlations (the eigenvalues of $\underset{\sim}{R}_{xx}^{-1}\underset{\sim}{R}_{xy}\underset{\sim}{R}_{yy}^{-1}\underset{\sim}{R}_{yx}$):

$$\Lambda = \prod_j (1 - \lambda_j), \quad j = 1...s.$$

Here the concept introduced earlier (Section 8.2) of a pooled within-groups dispersion matrix is generalized to that of an error dispersion matrix, where each $(1 - \lambda_j)$ is a measure of error variance. As the error variances are small, so Λ is small and the two data sets in question are highly correlated (Green, 1978).

The null hypothesis of zero correlation between data sets ($\underset{\sim}{R} = 0$), that is, of zero population canonical correlations, may be tested using Bartlett's χ^2 approximation of Λ,

$$\chi^2 = -[(n - 1) - 0.5(p + q + 1)]\ln\Lambda,$$

with pq degrees of freedom.

Such a test of the joint nullity of all s canonical correlations, if it leads to rejection of the null

hypothesis, may be followed by succesive tests of the joint nullity of the smaller s - k canonical correlations, k = 2...s, in order to obtain an indication of the dimensionality of the relationship between the data sets (Gittins, 1979).

$$\Lambda_k = \prod_i (1 - \lambda_i), \quad i = k...s, \quad k = 2...s.$$

The corresponding X^2 value is compared with tabulated values for $(p - k + i)(q - k + 1)$ degrees of freedom.

Alternatively, the significance of the squared canonical correlations may be tested by means of Roy's union-intersection test (Gittins, 1979). Gittins describes certain advantages that this test has over the partitioned Λ tests. The test involves comparing the eigenvalues of $\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}$ with the critical values of Roy's largest-root criterion, $x_\alpha(s,m,n)$, tabulated by Pillai (1960) and graphed by Heck (1960; Beyer, 1968). The null hypothesis of zero population canonical correlations referred to above is rejected if $\lambda_1 > x_\alpha(s,m,n)$, where $s = \min(p,q)$, $M = (p - q - 1)/2$, $n = (N - p - q - 2)/2$, and N is the sample size. Successive eigenvalues (λ_k, k = 2...s) may be tested, after rejecting the initial null hypothesis, by adjusting s for each one, as $s = \min(p - k + 1, q - k + 1)$.

Structure correlations (Cooley & Lohnes, 1971) or coefficients (Gittins, 1979) describe the linear

relationships between canonical variates and the original descriptors. These may be of two kinds (Gittins, 1979):

<u>intraset correlations</u> $\underset{\sim}{G} = \underset{\sim yy\sim}{R}\underset{\sim}{A}$,

$\underset{\sim}{H} = \underset{\sim xx\sim}{R}\underset{\sim}{V}$, and

<u>interset correlations</u> $\underset{\sim}{GG} = \underset{\sim xy\sim}{R}\underset{\sim}{A}$,

$\underset{\sim}{HH} = \underset{\sim yx\sim}{R}\underset{\sim}{V}$.

The former describe the contribution each descriptor makes to each canonical variate of the same data set, whereas the latter describe the relationship between descriptors from one set and the canonical variates of the other.

<u>Descriptor communalities</u> (Gittins, 1979) describe the proportion of a descriptor's variance accounted for by the r canonical variates retained on the basis of the significance tests described above. As above, these are intraset communalities

$$H_{wi} = \sum_k g_{ik}^2, \quad i = 1...p, \quad k = 1...r,$$

$$H_{wj} = \sum_k h_{jk}^2, \quad j = 1...q, \quad k = 1...r,$$

and interset communalities

$$H_{bi} = \sum_k gg_{ik}^2, \quad i = 1...p, \quad k = 1...r,$$

$$H_{bj} = \sum_k hh_{jk}^2, \quad j = 1...q, \quad k = 1...r.$$

Finally, the proportion of the variance of one data set explained by a canonical variate of the other set, given the availability of the second data set, is described as redundancy. This is an asymmetric measure of accounted-for variance, in contrast to the canonical correlation which is symmetric, and whose square is a measure of shared variance between the two data sets.

$$Rd_{T_k|U_k} = \lambda_k (\underset{\sim}{g}_k' \underset{\sim}{g}_k)/p,$$

$$Rd_{U_k|T_k} = \lambda_k (\underset{\sim}{h}_k' \underset{\sim}{h}_k)/q,$$

for k = 1...s, and where U_k and T_k are the canonical variates for the two data sets.

8.4.1 60 OTU Sample

For the 60 OTU subsample both leaf, and flower and fruit data were available. Accordingly, CCA was used to describe the relationship between these two data sets. All six leaf descriptors in Table 3.3 were employed, plus the six flower and fruit descriptors whose selection is described in Section 7.4. The analysis was implemented using both Cooley & Lohnes' program CANON (Cooley & Lohnes, 1971) and program 6M of the Biomedical Computer Programs P-Series (BMDP; Dixon & Brown, 1979). Additional statistics (Green, 1978; Gittins, 1979) were calculated using a MAXBASIC program written for the purpose.

The correlation structure for the sample (**R**) is shown in Table 8.4, for all the descriptors used. The canonical correlations for the first two pairs of canonical variates are high, and the results of the significance tests used (Table 8.5) suggest that only these two need to be retained for a description of the covariance of leaf, and flower and fruit data in this sample. Plots of OTUs on the first two canonical variates for each data set (Fig. 8.1) show a somewhat greater tendency for individual topodeme samples to remain distinct from one another in the flower and fruit domain than in the leaf domain. This is not surprising in view of the high degree of structure in the former data set already found to correspond to the topodeme structure of the sample (Section 6.2, 6.3).

Since each of the canonical variate pairs are pairs of linear compounds of the original descriptors that maximize the correlation of the two sets of descriptors, it is interesting to note the degree to which this maximization results in topodeme sample groupings of the OTUs even in the leaf domain (compare Fig. 6.10, Fig. 8.1). Intraset and interset correlations between canonical variates and descriptors are indicated in Table 8.7 and Fig. 8.2. Here, these are of value in interpreting the positions of topodeme samples in Fig. 8.1. In ecological applications of CCA it may be possible to attribute interset correlations to causal

relationships, for example between the physical environment and vegetation (e.g. Eldred & Maun, 1982). Such relationships are unlikely to exist between data sets obtained from different parts of the same individuals, however.

The two pairs of retained canonical variates differ slightly in the extent to which they summarize both the corresponding data sets and the respective other data set. The first two flower and fruit canonical variates (T_1, T_2) account for 69.4% of the variance of the flower and fruit descriptors (Table 8.7). The first two leaf canonical variates (U_1, U_2) account for only 51.4% of the variance of the leaf descriptors, however (Table 8.7). Similarly, whereas given the leaf data, 47.4% of the flower and fruit variance is explained by U_1 and U_2, only 35.3% of the leaf variance is explained by T_1 and T_2, given the flower and fruit data (Table 8.7). This suggests that given the canonical correlation of the two data sets, the leaf data is less completely summarized in only two dimensions than is the flower and fruit data.

As noted already (Section 6.2.1, 7.5) some descriptors have appreciably larger variances than others. Here, two of the flower descriptors (STYL, TCAL) have large coefficients of variation compared to those of the other ten descriptors in the analysis (Table 8.4), and are strongly correlated

with the two retained pairs of canonical variates (Table 8.7). The question arises whether the dimensionality of the data found here (Table 8.5) is a function of the inclusion of these descriptors in the analyses. However, repeating the analysis (not shown) without these descriptors still results in rejecting the null hypothesis of the independence of the two data sets.

Figure 8.1 Canonical correlation analysis of the 60 OTU subsample. Ordination is in the space defined by the first two canonical variates of the flower and fruit (T_1, T_2) and leaf (U_1, U_2) domains. See text for details. Vectors represent contribution of each descriptor to the scatter of OTUs in terms of the corresponding canonical weights (Table 8.6). The 10-stamen morphotype of C. crus-galli is represented by solid symbols, the 20-stamen one by open ones, and C. punctata (T7, Site 5) by the letter p. Individual sites are coded by geometric shape: squares (Site 1; T1, T2); circles (Site 2; T4, T5); triangles (Site 3; T3) and stars (Site 4; T6). Open stars indicate OTUs of C. ?grandis (Site 4). See Table 3.1 and 3.3 for explanation of descriptor abbreviations.

CANONICAL CORRELATION ANALYSIS,

Raw OTU means.

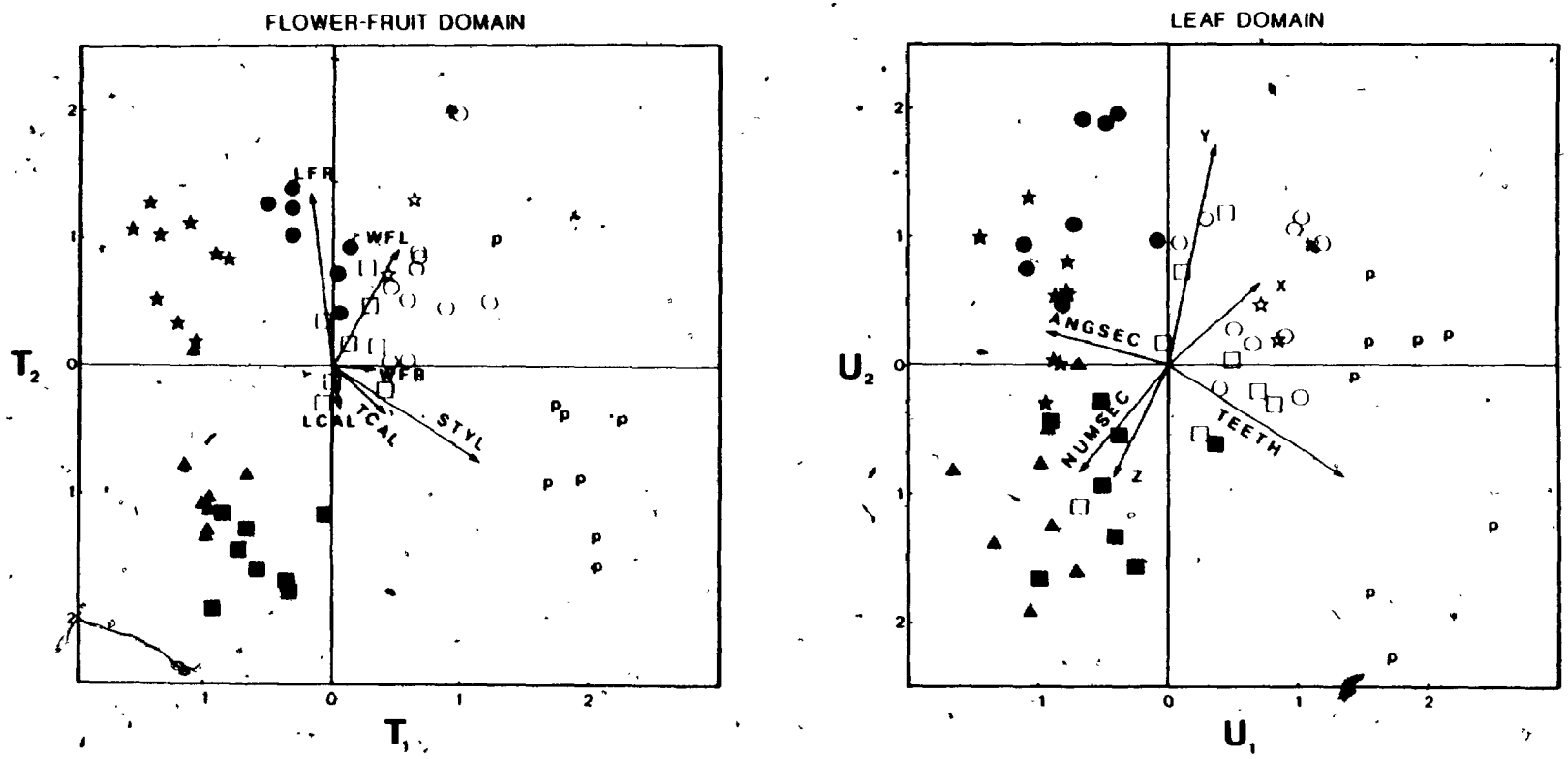


Figure 8.2 Canonical analysis of the relationships between 6 flower and fruit descriptors (Table 3.1) and 6 leaf descriptors (Table 3.3) for the 60 OTU subsample. For the ordination depicted in Figure 8.1, vectors represent intraset (solid) and interset (dashed) correlations between the canonical variates shown and the descriptors indicated (see tables indicated for explanation of descriptor abbreviations; see Table 8.7 for correlations).

CANONICAL CORRELATION ANALYSIS,

structure coefficients.

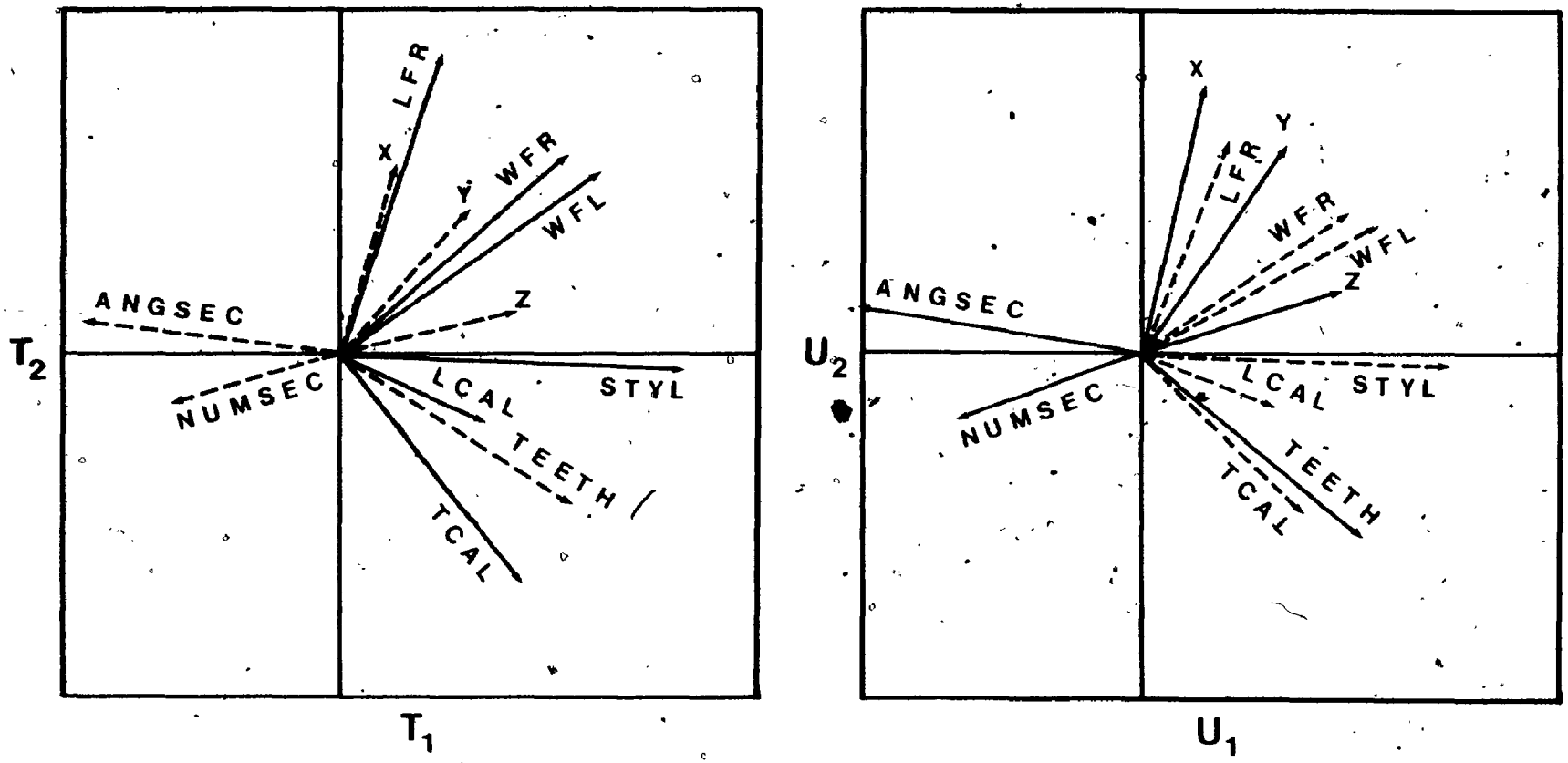


Table 8.4 Correlation matrix R for 60 OTU subsample and 6 flower and fruit descriptors (Table 3.1) plus 6 leaf descriptors (Table 3.3), calculated from raw OTU means. Entries along the principal diagonal are descriptor coefficients of variation. See tables indicated for explanation of descriptor abbreviations.

	STYL	TCAL	WFL	LCAL	LFR	WFR	X	Y	Z	NUMSEC	ANGSEC	TEETH
STYL	0.35											
TCAL	0.44	0.46										
WFL	0.62	0.04	0.11									
LCAL	0.44	0.17	0.15	0.11								
LFR	0.29	-0.43	0.47	0.15	0.10							
WFR	0.59	-0.09	0.63	0.25	0.76	0.13						
X	0.16	-0.26	0.35	-0.12	0.54	0.39	0.14					
Y	0.29	-0.06	0.51	0.27	0.47	0.49	0.56	0.12				
Z	0.46	0.17	0.48	0.22	0.21	0.37	0.49	0.62	0.12			
NUMSEC	-0.47	-0.10	-0.45	0.03	-0.16	-0.34	-0.07	0.16	-0.21	0.13		
ANGSEC	-0.69	-0.41	-0.53	-0.25	-0.06	-0.37	-0.05	-0.07	-0.55	0.67	0.14	
TEETH	0.55	0.56	0.21	0.45	-0.12	0.19	-0.35	0.24	0.21	0.13	-0.26	0.17

Table 8.5 Canonical analysis of the relationships between 6 flower and fruit descriptors (Table 3.1) and 6 leaf descriptors (Table 3.3) for the 60 OTU subsample. Canonical correlation coefficients and related tests of significance. Critical values of $x_\alpha(s,m,n)$ are taken from tables in Pillai (1960) and charts in Beyer (1968).

(a) Squared canonical correlation coefficients and their significance. The approximate critical value of Roy's largest-root criterion $x_\alpha(s,m,n)$ is given for $\alpha = 0.05$ and $\alpha = 0.01$ ($m = -0.5$; $n = 23$). See text for details.

k	s	$\lambda_k^{\frac{1}{2}}$	λ_k	$x_{.05}(s,m,n)$	$x_{.01}(s,m,n)$	p
1	6	0.906	0.8201	< 0.453	< 0.511	< 0.01
2	5	0.706	0.4989	0.345	0.405	< 0.01
3	4	0.541	0.2923	0.295	0.355	> 0.05
4	3	0.317	0.1002	0.235	0.290	> 0.05
5	2	0.105	0.0111	0.160	0.220	> 0.05
6	1	0.015	0.0002			

Table 8.5 Cont.

(b) Bartlett's approximate test of the joint nullity of the smallest s - k canonical correlations.

k	Canonical Correlations	X^2	d.f.	p
0	1,2,3,4,5,6	150.63	36	< 0.00001
1	2,3,4,5,6	60.57	25	0.00009
2	3,4,5,6	24.29	16	0.08334
3	4,5,6	6.14	9	0.72606
4	5,6	0.60	4	0.96357
5	6	0.01	1	0.91467

Table 8.6 Canonical analysis of the relationships between 6 flower and fruit descriptors (Table 3.1) and 6 leaf descriptors (Table 3.3) for the 60 OTU subsample. Canonical weights for standardized descriptors (zero mean, unit standard deviation) for each domain. See tables indicated for explanation of descriptor abbreviations. See also Figure 8.1.

(a) Flower - fruit domain canonical variates:

	T_1	T_2	T_3	T_4	T_5	T_6
STYL	0.607	-0.378	1.317	-0.267	-0.100	-1.030
TCAL	0.218	-0.184	-0.267	0.856	-0.376	1.070
WFL	0.271	0.477	-0.807	-0.667	-0.772	0.331
LCAL	0.030	-0.169	-0.959	0.333	-0.014	-0.483
LFR	-0.081	0.686	0.408	1.334	-1.006	0.060
WFR	0.177	-0.007	-0.409	-0.293	1.784	0.594

(b) Leaf domain canonical variates:

	U_1	U_2	U_3	U_4	U_5	U_6
X	0.361	0.325	1.255	0.862	-0.249	-0.028
Y	0.179	0.858	-1.250	-0.051	0.217	0.603
Z	-0.211	-0.432	-0.063	-0.625	0.021	-1.576
NUMSEC	-0.349	-0.420	-0.369	0.632	-1.233	0.194
ANGSEC	-0.472	0.134	0.320	-0.106	1.411	-1.109
TEETH	0.678	-0.433	0.609	0.749	0.666	-0.226

Table 8.7 Canonical analysis of the relationships between 6 flower and fruit descriptors (Table 3.1) and 6 leaf descriptors (Table 3.3) for the 60 OTU subsample. Correlations between the original descriptors and the first three canonical variates of each domain (see Fig. 8.2).

(a) Flower - fruit domain:

	T_1	T_2	T_3	H(1)	U_1	U_2	U_3	H(2)
STYL	0.964	-0.042	0.157	0.931	0.873	-0.030	0.086	0.763
TCAL	0.518	-0.655	-0.027	0.697	0.469	-0.462	-0.014	0.434
WFL	0.734	0.528	-0.210	0.818	0.665	0.373	-0.113	0.581
LCAL	0.406	-0.199	-0.588	0.204	0.367	-0.141	-0.318	0.155
LFR	0.265	0.849	0.080	0.791	0.240	0.600	0.045	0.417
WFR	0.634	0.565	-0.047	0.721	0.573	0.400	-0.026	0.489
Variance(3)	39.6	29.8	7.0	100.0(4)				
Redundancy(5)					32.5	14.9	2.1	50.3(6)

Table 8.7 Cont.

(b) Leaf domain:

	U_1	U_2	U_3	H(1)	T_1	T_2	T_3	H(2)
X	0.169	0.766	0.317	0.615	0.153	0.541	0.171	0.317
Y	0.399	0.594	-0.515	0.512	0.361	0.419	-0.279	0.306
Z	0.555	0.178	-0.188	0.340	0.502	0.125	-0.102	0.268
NUMSEC	-0.526	-0.186	-0.334	0.311	-0.477	-0.131	-0.180	0.244
ANGSEC	-0.796	0.124	-0.023	0.649	-0.720	0.088	-0.011	0.527
TEETH	0.627	-0.519	-0.278	0.662	0.568	-0.367	-0.150	0.457
Variance(3)	29.9	21.5	9.8	100.0(4)				
Redundancy(5)					24.6	10.7	2.9	40.1(6)

Table 8.7 Cont.

(1) Intraset descriptor communalities for the first two canonical variates.

(2) Interset descriptor communalities for the first two canonical variates.

(3) Percent domain variance explained by the corresponding canonical variate.

(4) Total percent domain variance explained by all its canonical variates.

(5) Percent domain variance explained by the corresponding canonical variate of the other domain (redundancy).

(6) Total percent domain variance explained by all the canonical variates of the other domain (total redundancy).

8.5 Canonical Variates Analysis

Canonical analysis may be used to explore a hypothesis concerning the a priori grouping of multivariate observations. The methodology of such an exploration may be developed from that of the multivariate analysis of variance (Section 8.2) or CCA (Section 8.4). With either approach the eigenstructure of an asymmetric matrix product is found and used to obtain linear compounds of the original descriptors (canonical variates) which maximally depict (in the sense of an ordination of data points) the hypothesized group structure. The eigenanalysis leads also to statistical tests of the grouping hypothesis. This methodology is referred to here as canonical variates analysis (CVA), following Gittins (1979) and others (e.g. Campbell, 1979; Campbell & Atchley, 1981).

The first of the approaches mentioned is based on the multivariate analysis of the variance, and is used by Cooley and Lohnes (1971) in their presentation of multiple group discriminant analysis, the extention of Fisher's two-group discriminant function. Following partition of the total variance of the sample into a pooled within-groups component $\underset{\sim}{W}$ and an among groups component $\underset{\sim}{A}$ (Section 7.1), the eigenstructure of $\underset{\sim}{W}^{-1}\underset{\sim}{A}$ is determined, resulting in a diagonal matrix $\underset{\sim}{D}$ of eigenvalues and a matrix $\underset{\sim}{K}$ of eigenvectors.

As described already (Section 8.2), the determinants of $\underset{\sim}{W}$ and of $\underset{\sim}{W} + \underset{\sim}{A}$ provide the basis for testing null hypotheses concerning the presence or absence of distinct groups. Tests of the significance of canonical variates are closely related, and are based on the equivalence of the eigenvalues of $\underset{\sim}{W}^{-1}\underset{\sim}{A}$ and of $\underset{\sim xx}{R}^{-1}\underset{\sim xy}{R}\ \underset{\sim yy}{R}^{-1}\underset{\sim yx}{R}$, described below. The eigenvectors of $\underset{\sim}{W}^{-1}\underset{\sim}{A}$ in $\underset{\sim}{K}$ may be rescaled in a number of ways (Green, 1978; Gittins, 1979; Pimentel, 1979) so as to obtain ordinations of the data with slightly different properties.

The second approach is due to Bartlett (1938), who established the formal equivalence of multiple group discriminant analysis and canonical correlation analysis in which one of the data sets represents the hypothesis of g groups by means of $q = g - 1$ binary-valued dummy variables. Accordingly, it can be shown (Green, 1978) that the eigenvalues of the corresponding matrix product $\underset{\sim xx}{R}^{-1}\underset{\sim xy}{R}\underset{\sim yy}{R}^{-1}\underset{\sim yx}{R}$, symbolized λ, and those of $\underset{\sim}{W}^{-1}\underset{\sim}{A}$, symbolized $^*\lambda$, are related as follows:

$$(\lambda)^{\frac{1}{2}} = (^*\lambda/(1 - {}^*\lambda))^{\frac{1}{2}}$$

and

$$^*\lambda = \lambda/(1 - \lambda).$$

As a result, the signficance tests for canonical correlation analysis (Section 8.4) may be adapted to testing the significance of the canonical variates obtained here.

If the eigenvectors of $\underset{\sim}{W}^{-1}\underset{\sim}{A}$ are scaled by prior standardization of the descriptors to zero mean and unit standard deviation across the entire sample they will be equivalent to the canonical weights derived from the eigenvectors of $\underset{\sim}{R}_{xx}^{-1}\underset{\sim}{R}_{xy}\underset{\sim}{R}_{yy}^{-1}\underset{\sim}{R}_{yx}$ for the observed set of p descriptors. This matrix of eigenvectors $\underset{\sim}{K}_t^*$ (in the notation of Green, 1978) is the one employed in calculating the structure coefficients described below.

An alternative scaling of the canonical weights (Green, 1978) is used for graphical representation of the results obtained in the analyses described below. Scores calculated directly from the eigenvectors of $\underset{\sim}{W}^{-1}\underset{\sim}{A}$ have sums of squares equal to unity on each canonical variate. As a result, scales on each axis differ, and the group dispersions in the space defined by these axes are roughly (hyper-) ellipsoidal. The eigenvectors may be normalized, however, so that on each canonical variate the average within-group variance is unity. This "spherizing" transformation ($\underset{\sim}{K} \rightarrow \underset{\sim}{K}_w$; Green, 1978; Gittins, 1979) of the canonical variate scores results in (hyper-) spherical group dispersions, and has the effect also of maximizing distances between group centroids (Gittins, 1979). Accordingly, in the space of any k canonical variates the approximate confidence region around a group mean vector (sample size = N) is circular (k = 2) or (hyper-) spherical ($k \geq 3$), with

radius equal to

$$(\chi^2_{[k]}/Ng)^{\frac{1}{2}}$$ (Gnanadesikan, 1977).

A matrix $\underset{\sim}{H}$ of correlations between canonical variates and the original descriptors is obtained from the matrix of rescaled eigenvectors K^*_t described above and the correlation matrix $\underset{\sim}{R}_t$ obtained from the SSCP matrix $\underset{\sim}{T}$, as $\underset{\sim}{H} = \underset{\sim}{R}_t \underset{\sim}{K}^*_t$. The sum of the squared elements of $\underset{\sim}{H}$ for each descriptor are described by Cooley and Lohnes (1971) as descriptor communalities, high values of which indicate a high degree of association between the descriptor in question and the discrimination of the groups hypothesized. However, these are of interest only when $(g - 1) < p$, since they are equal to 1.00 otherwise (g - the number of groups; p = the number of observed descriptors).

The analyses described below (Section 8.5.1, 8.5.2.1 - 8.5.2.4) have been implemented using the approach described by Cooley and Lohnes (1971), and their program DSCRM. This program uses as input the $\underset{\sim}{W}$ and $\underset{\sim}{T}$ matrices output by program MANOVA. Its output has been modified by R. Gittins so as to include the canonical variate scores for both of the scalings of the canonical weights described above. These analyses are discussed in terms of the two data sets available, leaf (Section 8.5.1) and flower and fruit (Section 8.5.2).

8.5.1 Leaf Data

Canonical variates analysis of the ranged leaf data (Table 3.3) available for the 60 OTU subsample (Table 3.2) was used to portray only the hypothesis of seven groups corresponding to the seven topodeme samples (Chapter 7; Section 8.2). As described earlier (Section 7.1) it was necessary to restrict analyses of group structure to the 58 crus-galli and punctata OTUs in this subsample, omitting two OTUs of C. ?grandis (Table 3.2) in order to have homogeneous groups of the maximum possible size.

The null hypothesis of the equality of the canonical roots (eigenvalues of $\underset{\sim}{R}_{xx}^{-1}\underset{\sim}{R}_{xy}\underset{\sim}{R}_{yy}^{-1}\underset{\sim}{R}_{yx}$) is rejected for the four largest ones by both Roy's union-intersection test and Bartlett's partitioned Λ test (Table 8.8b, c). The corresponding eigenvalues of $\underset{\sim}{W}^{-1}\underset{\sim}{A}$ account for 58.4, 28.4, 8.3 and 4.4% of the trace of $\underset{\sim}{W}^{-1}\underset{\sim}{A}$ respectively.

Although in terms of the canonical weights (Fig. 8.3) there is a shape contrast between broad versus narrow leaves (T3 versus T1), this contrast is less marked in terms of the correlations between canonical variates and descriptors (Table 8.8d; compare also Fig. 8.1, 8.2). The 10- and 20-stamen morphotypes of C. crus-galli are segregated from each other for the most part (Fig. 8.3), with the latter appearing to be intermediate between the 10-stamen

C. crus-galli OTUs and those of C. punctata, particularly with respect to leaf margination (TEETH). As seen earlier (Fig. 6.10) crus-galli and punctata are contrasted particularly with respect to the two multistate descriptors of the secondary venation, NUMSEC and ANGSEC (Table 3.3). In general, the leaves of C. crus-galli sensu lato are characterized by more secondary veins than are typical of the leaves of C. punctata. The secondary veins of the leaves of the latter species are larger in diameter and more conspicuous, however. In addition, they are also more strongly ascending (ANGSEC is usually smaller) than is the case with crus-galli, especially in the center of the leaf where ANGSEC is measured (compare Fig. 8.10d with the remainder of the figure and with Fig. 8.9; see also Table 3.4).

The analysis described here is based in part on the three leaf dimensions measured (X, Y, Z; Table 3.3). Although shape contrasts are evident in the disposition of OTUs in Fig. 8.3, the use of shape descriptors that are ratios of such dimensions has also been proposed (Byatt, 1975; Gostyńska-Jakuszewska, 1975; Sinnott, 1978; Sinnott & Phipps, 1983). It is of interest to note that repeating the analysis described above using instead of X, Y and Z two log-transformed ratios (Mosimann & James, 1979), narrowness, defined as (logY - Log(X + Z)), and lanceolateness, defined

as (logX - logZ), discrimination is actually lost. Using the raw descriptors $\Lambda = 0.0023$ (Table 8.8a), whereas the analysis based on NUMSEC, ANGSEC, TEETH, narrowness and lanceolateness (not shown) results in $\Lambda = 0.0047$. Thus, quite apart from the disadvantages of ratio descriptors described by Atchley, Gaskins, and Anderson (1976), it appears that in this case at least analyses based on the original dimensions adequately recover the contrasts expressed by their ratios.

Figure 8.3 Canonical variates analysis. The 58 OTU subsample in the space defined by the first two canonical variates calculated from pooled within- and among-groups dispersion matrices for 6 descriptors of short shoot terminal leaves (Table 3.3). Canonical variates are scaled so that the mean within-group variance weighted by sample size along each axis is unity. Dashed circles indicate location of topodeme sample centroids (T1 - T7), and approximate 99% confidence regions around these points. Vectors represent the contribution of the descriptor indicated to the scatter of the OTUs in terms of its canonical weights. See Figure 8.1 for explanation of symbols. See text and Table 8.8 for details.

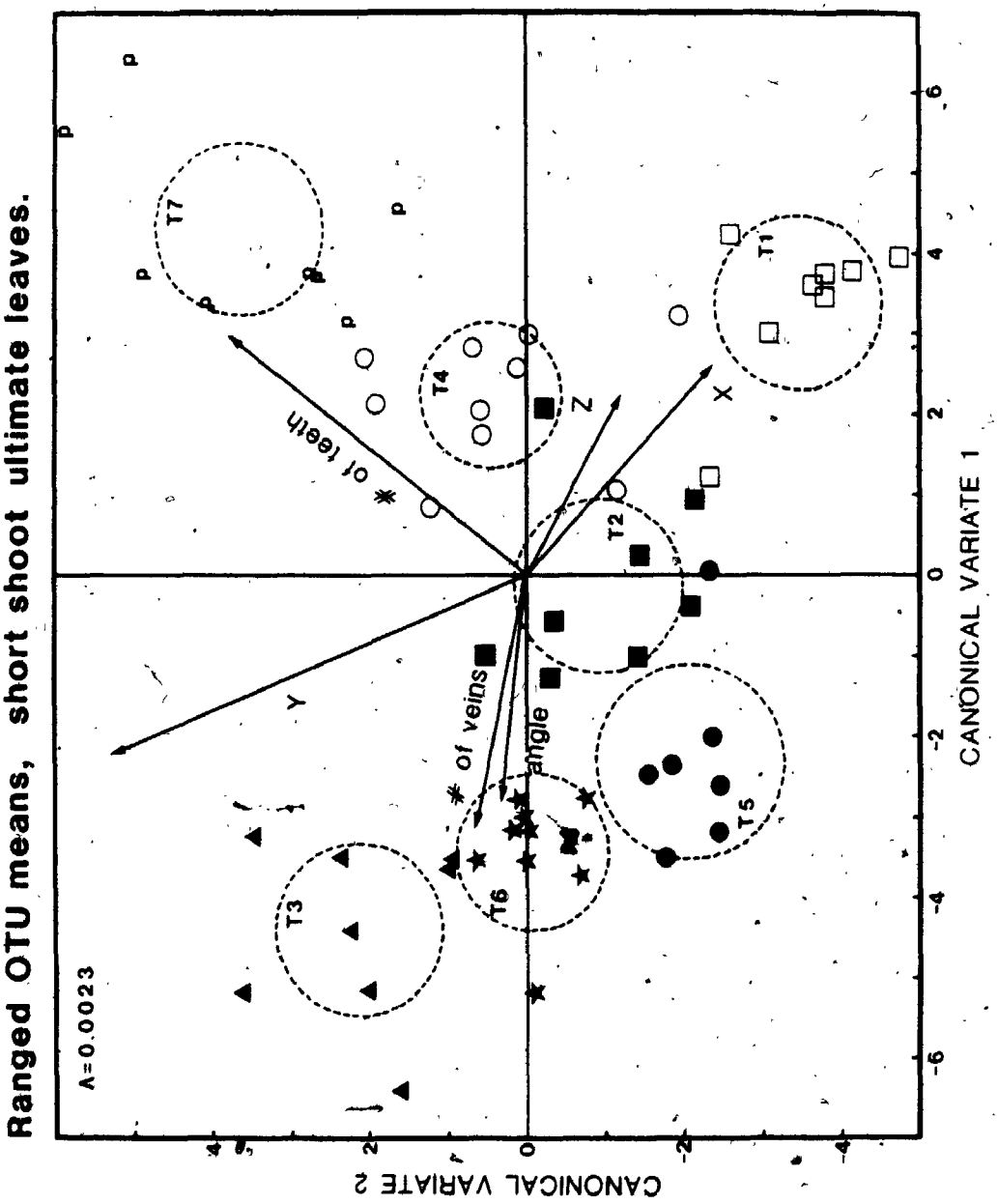

Ranged OTU means, short shoot ultimate leaves.
A=0.0023
T1
T2
T3
T4
T5
T6
T7
X
Y
Z
* of teeth
* of veins
angle
CANONICAL VARIATE 1
CANONICAL VARIATE 2

Table 8.8 Canonical variates analysis of the 58 OTU subsample using 6 ranged leaf descriptors (Table 3.3). See text for details. See also Figure 8.3. Groups are the seven topodeme samples (T1 - T7).

(a) Wilks' $\Lambda = 0.0023$

$n_1 = 36$ $n_2 = 204$ $F_s = 16.95$ $p << 0.001$

(b) Canonical correlation coefficients and their squares (eigenvalues of $R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}$ for leaf data and 6 binary valued dummy variables for group affiliation). Approximate critical values of Roy's largest-root criterion $x_\alpha(s,m,n)$ are given for $\alpha = 0.05$ and $\alpha = 0.01$ ($m = -0.5$, $n = 22$), taken from tables in Pillai (1960) and charts in Beyer (1968).

k	s	$\lambda_k^{\frac{1}{2}}$	λ_k	$x_{.05}(s,m,n)$	$x_{.01}(s,m,n)$	p
1	6	0.959	0.919	< 0.453	< 0.511	< 0.01
2	5	0.920	0.847	0.360	0.420	< 0.01
3	4	0.787	0.619	0.305	0.365	< 0.01
4	3	0.680	0.462	0.240	0.300	< 0.01
5	2	0.282	0.080	0.170	0.230	> 0.05
6	1	0.078	0.006			

Table 8.8 Cont.

(c) Bartlett's approximate test of the joint nullity of the smallest s - k canonical correlations.

k	Canonical Correlations	X^2	d.f.	p
0	1,2,3,4,5,6	306.32	36	< 0.005
1	2,3,4,5,6	179.75	25	< 0.005
2	3,4,5,6	84.52	16	< 0.005
3	4,5,6	35.78	9	< 0.005
4	5,6	4.50	4	> 0.100
5	6	0.31	1	> 0.500

(d) Correlations between canonical variates (I - VI) and the original descriptors.

	I	II	III	IV	V	VI
X	0.216	-0.391	-0.705	-0.347	0.313	0.293
Y	0.184	0.392	-0.641	0.090	0.507	0.370
Z	0.659	0.048	-0.037	-0.139	0.660	0.329
NUMSEC	-0.721	0.409	0.049	-0.360	0.124	0.407
ANGSEC	0.921	0.060	-0.179	-0.172	0.171	-0.240
TEETH	0.351	0.898	0.122	-0.198	-0.094	-0.090

8.5.2 Flower and Fruit Data

Canonical analyses of the flower and fruit data utilized six descriptors (STYL, TCAL, WFL, LCAL, LFR, WFR) whose selection was described in Section 7.3 Since data for the flower and fruit descriptors were available for the entire 160 OTU sample, CVA was used to examine four different samples of OTUs. In the first three cases the groups hypothesized are the topodeme samples. The first analysis (Section 8.5.2.1) examines only the 93 randomly sampled crus-galli OTUs. The next two (Section 8.5.2.2) deal with the crus-galli and punctata topodeme samples together, with respect first to the 58 OTU subsample (Table 3.2) and then to the total of 111 randomly sampled OTUs (excluding as before the two individuals of C. ?grandis). Finally the total sample of 160 OTUs is divided into taxonomic groups and CVA is used as a means of ordinating the sample somewhat more appropriate than the PCA employed earlier (Section 6.2.1).

8.5.2.1 93 OTU Sample

In Chapter 7 it was found that the presence of the punctata sample resulted in increased group covariance heterogeneity (Table 7.2). CVA of the six crus-galli topodeme samples by themselves was done in order to observe the effect on the ordination of the absence of the punctata

topodeme sample. The ordination of group centroids is presented in Fig. 8.6a.

The results of both Bartlett's partitioned test and Roy's union-intersection test indicate that the first four canonical roots are significant (Table 8.9b,c). The corresponding roots of $\underset{\sim}{W}^{-1}\underset{\sim}{A}$ account for 72.0, 17.4, 7.9 and 2.6% of the trace of $\underset{\sim}{W}^{-1}\underset{\sim}{A}$, respectively.

The correlations between canonical variates and the descriptors, and the descriptor communalities, suggest that while topodeme sample discrimination involves all six descriptors, those descriptors most highly correlated with stamen number such as style number (STYL), fruit width (WFR) and especially flower width (WFL) are particularly important (Table 8.9d). As a result, the principal split in the sample is according to stamen number, with only T5 occupying an intermediate position (Fig. 8.6a).

8.5.2.2 58 OTU and 111 Sample

CVA of all seven topodeme samples together resulted in finding five significant canonical roots for both the 58 OTU subsample (Table 8.10b, c) and the total sample of 111 OTUs (Table 8.11c,d). These accounted for 79.4, 12.8, 4.8, 2.4 and 0.7%, and 70.8, 16.9, 7.6, 3.7 and 0.9% of the trace of $\underset{\sim}{W}^{-1}\underset{\sim}{A}$ in each case, respectively.

As with the analysis of the leaf data (Fig. 8.3) the topodeme samples of the 20-stamen morphotype of C. crus-galli (T1, T4) are intermediate in position between those of the 10-stamen morphotype (T2, T3, T6) and that of C. punctata (T7; Fig. 8.4). The high dimensionality of the data is indicated by the way in which although samples T1 and T5 overlap in the plane defined by the first two canonical variates of both analyses, they are well separated along the third axis in each case.

The correlations between canonical variates and descriptors in the two analyses (Table 8.10d, 8.11e) show how, even more than in the preceding analysis of the crus-galli topodeme samples by themselves, the descriptors highly correlated with stamen number (STYL, WFL, WFR) are very important in discriminating C. punctata from 10- and 20-stamen C. crus-galli, as well as the two crus-galli morphotypes from each other. The same is true also for the canonical weights (Fig. 8.4), although less so for the smaller sample (not shown).

The ordination of the 111 OTU sample (Fig. 8.4) also depicts the contrast in topodeme sample variability discussed earlier (Section 7.3; Table 7.2, 7.3), between T7 (C. punctata) and the topodeme samples of 10- and 20-stamen C. crus-galli (T1 - T6).

8.5.2.3 155 OTU Sample

CVA was used to ordinate the entire 160 OTU sample (Fig. 11.1), based on five groups of OTUs, according to taxonomic affinity: the common 10- and 20-stamen morphotypes of C. crus-galli sensu lato; C. sp. aff. C. bushii; C. ?grandis; and C. punctata. This breakdown of the sample required omitting five OTUs from the analysis, since they could not be assigned to any one of these groups (two intermediate individuals from Site 4, OTUs 696 and 697, and the three OTUs belonging to section Macracanthae, OTUs 761, 762 and 769). After obtaining canonical weights for the spherizing transformation used in the ordination, however, the canonical variate scores for these OTUs were calculated so as to be able to plot them with the rest of the sample.

In this analysis all four canonical roots were found to be significant (Table 8.12b, c). These accounted for 49.4, 32.2, 15.4 and 2.9% of the trace of $W^{-1}A$ respectively. Again, 20-stamen C. crus-galli comes out between the 10-stamen morphotype and C. punctata (Fig. 11.1), although separated from both to some extent along the third canonical variate. In the opposite direction from the 20-stamen morphotype along the third canonical variate, and away to either side of the crus-galli - punctata trend are found the OTUs of C. sp. aff. C. bushii and C. ?grandis. The

location of the former appears to be a function particularly of its calyx lobe margination (TCAL; Fig. 11.1; Table 8.12d). This descriptor is largely responsible for the position of the Macracanthae OTUs as well. The OTUs of C. ?grandis appear to be separated from the rest of the sample principally on the basis of high values for calyx lobe length (LCAL) and fruit size (LFR, WFR).

Figure 8.4 Canonical variates analysis. 111 OTU sample in the space defined by the first two canonical variates calculated from the pooled within- and among-groups dispersion matrices for 6 flower and fruit descriptors (Table 3.1). Canonical variates are scaled so that the mean within-group variance weighted by sample size on each axis is unity. Dashed circles indicate location of topodeme sample centroids (T1 - T7), and approximate 99% confidence regions around these points. Vectors represent the contribution of the descriptor indicated to the scatter of OTUs in terms of its canonical weights. See Figure 8.1 for explanation of symbols. See text and Table 8.11 for details.

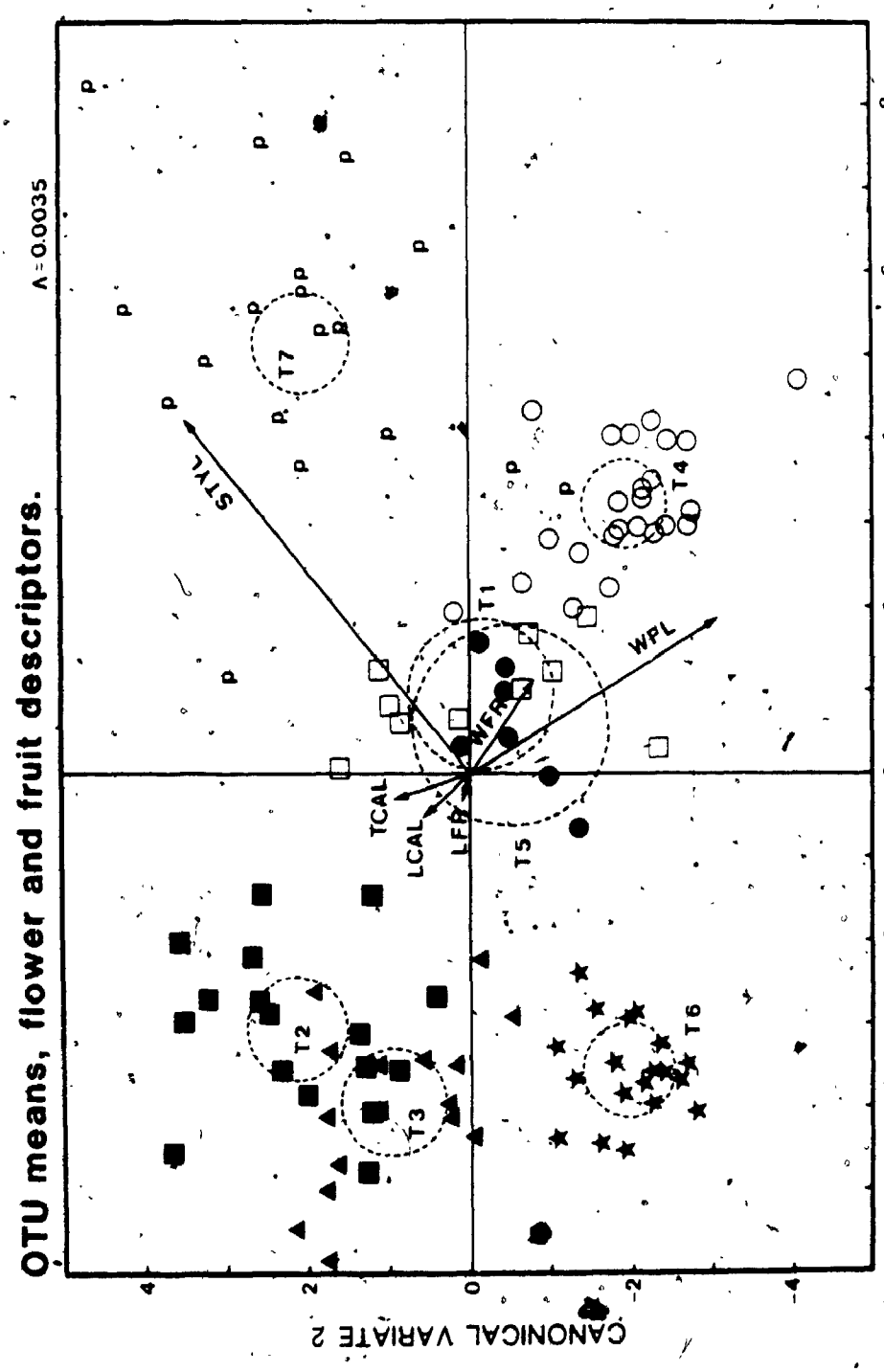

OTU means, flower and fruit descriptors.
Λ=0.0035
CANONICAL VARIATE 2
CANONICAL VARIATE 1
STVL
WFR
WFL
TCAL
LCAL
LFR
T1
T2
T3
T4
T5
T6
T7
p

Table 8.9 Canonical variates analysis of six topodeme samples of C. crus-galli s. l. (T1 - T6) using 6 ranged flower and fruit descriptors (Table 3.1). See text for details.

(a) Wilks' $\Lambda = 0.0025$

$n_1 = 30$ $n_2 = 330$ $F_s = 38.13$ $p << 0.001$

(b) Canonical correlation coefficients and their squares (eigenvalues of $R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}$ for flower and fruit data and 5 binary-valued dummy variables for group affiliation). Approximate critical values of Roy's largest-root criterion $x_\alpha(s,m,n)$ are given for $\alpha = 0.05$ and $\alpha = 0.01$ ($m = 0$, $n = 40$), taken from tables in Pillai (1960) and charts in Beyer (1968).

k	s	$\lambda_k^{\frac{1}{2}}$	λ_k	$x_{.05}(s,m,n)$	$x_{.01}(s,m,n)$	p
1	5	0.971	0.943	0.245	0.285	< 0.01
2	4	0.895	0.800	0.205	0.245	< 0.01
3	3	0.804	0.646	0.165	0.205	< 0.01
4	2	0.610	0.372	0.120	0.160	< 0.01
5	1	0.079	0.006			> 0.75(1)

(1) From the F distribution:

$F_s = ((n+1)/(m+1))(\lambda_k/(1-\lambda_k))$, $n_1 = 2m+2$, $n_2 = 2n+2$ (Morrison, 1976). $F_s = 0.25 < F_{.75[2,82]} = 0.289$.

Table 8.9 Cont.

(c) Bartlett's approximate test of the joint nullity of the smallest s - k canonical correlations.

k	Canonical Correlations	X^2	d.f.	p
0	1,2,3,4,5	514.85	30	< 0.005
1	2,3,4,5	268.42	20	< 0.005
2	3,4,5	129.89	12	< 0.005
3	4,5	40.54	6	< 0.005
4	5	0.54	2	> 0.500

(d) Correlations between canonical variates (I - V) and the original descriptors.

	I	II	III	IV	V	H(2)
STYL	0.924	0.301	0.071	-0.179	-0.077	0.988
TCAL	0.088	0.812	0.054	0.307	-0.283	0.844
WFL	0.872	-0.380	0.076	0.254	-0.159	1.000
LCAL	-0.004	0.224	-0.453	0.327	-0.646	0.745
LFR	0.410	-0.661	-0.128	-0.472	-0.238	0.901
WFR	0.826	-0.225	-0.451	-0.190	-0.062	0.976

(2) Descriptor communalities (Cooley & Lohnes, 1971).

Table 8.10 Canonical variates analysis of the 58 OTU subsample using 6 ranged flower and fruit descriptors (Table 3.1). See text for details. Groups are the seven topodeme samples (T1 - T7).

(a) Wilks' $\Lambda = 0.0010$

$n_1 = 36$ $n_2 = 204$ $F_s = 21.69$ $p << 0.001$

(b) Canonical correlation coefficients and their squares (eigenvalues of $R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}$ for flower and fruit data and 6 binary-valued dummy variables for group affiliation). Approximate critical values of Roy's largest-root criterion $x_\alpha(s,m,n)$ are given for $\alpha = 0.05$ and $\alpha = 0.01$ ($m = -0.5$, $n = 22$), taken from tables in Pillai (1960) and charts in Beyer (1968).

k	s	$\lambda_k^{\frac{1}{2}}$	λ_k	$x_{.05}(s,m,n)$	$x_{.01}(s,m,n)$	p
1	6	0.983	0.966	< 0.453	< 0.511	< 0.01
2	5	0.905	0.819	0.360	0.420	< 0.01
3	4	0.793	0.629	0.305	0.365	< 0.01
4	3	0.674	0.455	0.240	0.300	< 0.01
5	2	0.447	0.200	0.170	0.230	< 0.05
6	1	0.023	0.001			> 0.75(1)

(1) From the F distribution:

$F_s = ((n+1)/(m+1))(\lambda_k/(1-\lambda_k))$, $n_1 = 2m+2$, $n_2 = 2n+2$

Table 8.10 Cont.

(c) Bartlett's approximate test of the joint nullity of the smallest s - k canonical correlations.

k	Canonical Correlations	X^2	d.f.	p
0	1,2,3,4,5,6	348.47	36	< 0.005
1	2,3,4,5,6	178.23	25	< 0.005
2	3,4,5,6	91.97	16	< 0.005
3	4,5,6	41.92	9	< 0.005
4	5,6	11.27	4	> 0.100
5	6	0.03	1	> 0.500

(d) Correlations between canonical variates (I - VI) and the original descriptors.

	I	II	III	IV	V	VI
STYL	0.958	0.255	-0.044	-0.087	-0.009	-0.092
TCAL	0.398	0.780	0.103	0.175	0.012	0.438
WFL	0.807	-0.428	-0.042	0.343	0.214	-0.026
LCAL	0.258	0.508	-0.350	0.065	0.701	-0.238
LFR	0.406	-0.706	-0.264	-0.401	0.203	0.256
WFR	0.678	-0.355	-0.606	-0.002	-0.119	0.178

Table 8.11 Canonical variates analysis of the 111 OTU sample using 6 ranged flower and fruit descriptors (Table 3.1). See text for details. See also Figure 8.4. Groups are the seven topodeme samples (T1 - T7).

(a) Among- and pooled within-groups dispersion matrices. Diagonal elements are variances of the ranged descriptors; off-diagonal elements are descriptor correlations.

	STYL	TCAL	WFL	LCAL	LFR	WFR	
	0.019						STYL
	0.512	0.009					TCAL
A =	0.513	0.050	0.026				WFL
	0.404	0.211	0.104	0.002			LCAL
	0.315	-0.061	0.348	0.027	0.006		LFR
	0.531	0.062	0.417	0.103	0.566	0.017	WFR
	0.002						STYL
	0.064	0.009					TCAL
W =	0.048	0.035	0.004				WFL
	0.043	0.080	0.059	0.006			LCAL
	-0.008	-0.011	0.029	-0.043	0.005		LFR
	0.035	0.010	0.043	-0.066	0.459	0.004	WFR

(b) Wilks' $\Lambda = 0.0035$

$n_1 = 36$ $n_2 = 437$ $F_8 = 31.92$ $p << 0.001$

Table 8.11 Cont.

(c) Canonical correlation coefficients and their squares (eigenvalues of $R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx}$ for flower and fruit data and 6 binary-valued dummy variables for group affiliation). Approximate critical values of Roy's largest-root criterion $x_\alpha(s,m,n)$ are given for $\alpha = 0.05$ and $\alpha = 0.01$ ($m = -0.5$, $n = 43.5$), taken from tables in Pillai (1960) and charts in Beyer (1968).

k	s	$\lambda_k^{\frac{1}{2}}$	λ_k	$x_{.05}(s,m,n)$	$x_{.01}(s,m,n)$	p
1	6	0.964	0.930	< 0.278	< 0.319	< 0.01
2	5	0.872	0.760	0.205	0.250	< 0.01
3	4	0.767	0.588	0.170	0.215	< 0.01
4	3	0.641	0.411	0.135	0.175	< 0.01
5	2	0.382	0.146	0.095	0.125	< 0.01
6	1	0.004	0.000			> 0.75(1)

(1) From the F distribution:

$F_s = ((n+1)/(m+1))(\lambda_k/(1-\lambda_k))$, $n_1 = 2m+2$, $n_2 = 2n+2$ (Morrison, 1976). $F_s = 0.00 < F_{.75[1,99]} = 0.102$.

Table 8.11 Cont.

(d) Bartlett's approximate test of the joint nullity of the smallest s - k canonical correlations.

k	Canonical Correlations	X^2	d.f.	p
0	1,2,3,4,5,6	585.51	36	< 0.005
1	2,3,4,5,6	310.49	25	< 0.005
2	3,4,5,6	162.78	16	< 0.005
3	4,5,6	71.01	9	< 0.005
4	5,6	16.28	4	< 0.005
5	6	0.00	1	> 0.995

(e) Correlations between canonical variates (I - VI) and the original descriptors.

	I	II	III	IV	V	VI
STYL	0.941	0.334	-0.036	0.006	0.007	0.034
TCAL	0.381	0.646	0.011	-0.280	0.368	-0.473
WFL	0.869	-0.438	-0.094	-0.096	0.186	-0.006
LCAL	0.255	0.268	0.251	0.082	0.833	0.315
LFR	0.477	-0.402	-0.068	0.715	0.011	-0.308
WFR	0.822	-0.255	0.393	0.254	-0.070	-0.192

Table 8.12 Canonical variates analysis of 155 OTUs in five groups using 6 ranged flower and fruit descriptors (Table 3.1). See text for details. See also Figure 11.1. Groups: 10-stamen C. crus-galli; 20-stamen C. crus-galli; C. sp. aff. C. bushii; C. ?grandis; and C. punctata. Null hypothesis of the equality of group covariance matrices for this sample rejected ($p < 0.001$).

(a) Wilks' $\Lambda = 0.0088$

$n_1 = 24$ $n_2 = 507$ $F_s = 60.93$ $p << 0.001$

(b) Canonical correlation coefficients and their squares (eigenvalues of $\underset{\sim}{R}_{xx}^{-1}\underset{\sim}{R}_{xy}\underset{\sim}{R}_{yy}^{-1}\underset{\sim}{R}_{yx}$ for flower and fruit data and 4 binary-valued dummy variables for group affiliation). Approximate critical values of Roy's largest-root criterion $x_\alpha(s,m,n)$ are given for $\alpha = 0.05$ and $\alpha = 0.01$ ($m = 0.5$, $n = 71.5$), taken from tables in Pillai (1960) and charts in Beyer (1968).

k	s	$\lambda_k^{\frac{1}{2}}$	λ_k	$x_{.05}(s,m,n)$	$x_{.01}(s,m,n)$	p
1	4	0.921	0.849	0.145	0.175	< 0.01
2	3	0.886	0.785	0.120	0.145	< 0.01
3	2	0.798	0.637	0.095	0.120	< 0.01
4	1	0.501	0.251			< 0.001(1)

Table 8.12 Cont.

(1) From the F distribution:

$F_s = ((n+1)/(m+1))(\lambda_k/(1-\lambda_k))$, $n_1 = 2m+2$, $n_2 = 2n+2$

(Morrison, 1976). $F_s = 16.20 > F_{.001[3,145]} = 5.726$.

(c) Bartlett's approximate test of the joint nullity of the smallest s - k canonical correlations.

k	Canonical Correlations	X^2	d.f.	p
0	1,2,3,4	702.94	24	< 0.005
1	2,3,4	422.16	15	< 0.005
2	3,4	193.60	8	< 0.005
3	4	43.01	3	< 0.005

(d) Correlations between canonical variates (I - IV) and the original descriptors.

	I	II	III	IV	H(2)
STYL	0.552	-0.794	0.021	-0.189	0.972
TCAL	0.825	0.331	0.420	-0.130	0.983
WFL	0.694	-0.414	-0.343	0.424	0.951
LCAL	0.235	-0.415	0.611	0.439	0.793
LFR	0.059	-0.548	0.154	0.413	0.499
WFR	0.375	-0.621	-0.038	0.502	0.779

(2) Descriptor communalities (Cooley & Lohnes, 1971).

8.6 Evaluation of CVA Results

The canonical variates methodology used here involves an assumption of the equality of group covariance structures that is violated in all but one of the analyses reported here (Chapter 7). It is of interest to determine the extent to which these analyses are affected by the violation of this assumption. Gower (1966) has demonstrated the equivalence of the configurations of group centroids obtained from PCoA of a matrix of Mahalanobis' generalized distances (Section 4.3.2) calculated using the pooled within-groups covariance matrix $\underset{\sim}{W}$, and from CVA. On this basis, Campbell (Campbell & Mahon, 1974) suggests evaluating the effect of heterogeneity of group covariance structures by comparing CVA results with those of PCoA of generalized distances calculated using various alternative covariance matrices. Those used here (following Campbell & Mahon, 1974) are the joint covariance matrices $\underset{\sim}{S}_{hi}$ obtained for each pair of topodeme samples h and i according to the methods described in Section 4.3.2.1 and 4.3.2.2. In addition, the matrix of information radii among topodeme samples was also calculated (from the descriptors state frequencies over all OTUs in each sample), and the sample centroids similarly ordinated using PCoA. These distance matrices are presented in Table 8.13.

Comparison of Fig. 8.5a-d and 8.6a-d indicates that while differences in covariance structure have in some cases a marked effect on the position of a topodeme sample centroid, nevertheless in each of the PCoAs (Fig. 8.5b-d, 8.6b-d, Table 8.14) all of the topodeme sample centroids remain distinct from one another, if not in the plane of the first two axes then at least in the space defined by the first three (not shown). Correlations among the four ordinations ((a) CVA; (b) PCoA, generalized distances calculated using the pooled covariance matrices for pairs of topodeme samples; (c) PCoA, generalized distances calculated using the common covariance matrices for pairs of topodeme samples; and (d) PCoA, information radii between samples) for the 93 OTU sample are as follows (significant correlations, $p < 0.05$, underlined):

(b) 0.692

(c) -0.143 -0.152

(d) 0.900 0.659 -0.283

(a) (b) (c).

Correlations among the four ordinations of the seven topodeme samples are as follows:

(b) 0.577

(c) 0.019 -0.131

(d) 0.791 0.654 -0.056

(a) (b) (c).

Correlations among the ordinations were calculated using the

method described in Section 4.5, representing each ordination by the topodeme sample scores on the four coordinate axes, and calculating correlations over the 24 or 28 entries. In the case of the CVA, only the scores on the first four canonical variates were used.

The effect of the differences in covariance structure among the topodeme samples, notable in the topodeme sample dispersion matrix determinants (Table 7.3) appears greatest with T5 and T6, samples with the smallest generalized variance for the descriptors considered here. This effect is particularly noticeable where there appears to be an interaction with sample size, in Fig. 8.5c and 8.6c. These ordinations are based on generalized distances calculated pairwise among topodeme samples using Anderson's (1958) solution to the Fisher-Behrens problem for unequal covariance matrices (Section 4.3.2.2). This produces a joint covariance matrix S_{hi} in which only min (N_h, N_i) OTUs are considered for both groups h and i, with sample sizes N_h and N_i respectively. Since T5 is the smallest topodeme sample available (N = 7), it is not surprising that its relative position should change markedly in Fig. 8.5c and 8.6c, since it results from distances calculated on the basis of only the first seven OTUs in each of the other topodeme samples.

Figure 8.5 Evaluation of the results of canonical variates analysis. Positions of the centroids of six C. crus-galli topodeme samples (T1 - T6) in the space defined by:

(a) the first two canonical variates obtained in the analysis described in Table 8.9.

(b) the first two principal coordinates obtained in the analysis described in Table 8.14a, based on generalized distances calculated pairwise among topodeme samples using the pooled covariance matrix (Table 8.13a).

(c) the first two principal coordinates obtained in the analysis described in Table 8.14b, based on the generalized distances calculated pairwise among topodeme samples using common covariance matrices (Anderson, 1958; Table 8.13b).

(d) the first two principal coordinates obtained in the analysis described in Table 8.14c, based on information radii calculated pairwise among the samples (Table 8.13c).

The scales of the axes in (a-c) are the same. See Figure 8.1 for explanation of the symbols used. See text for details.

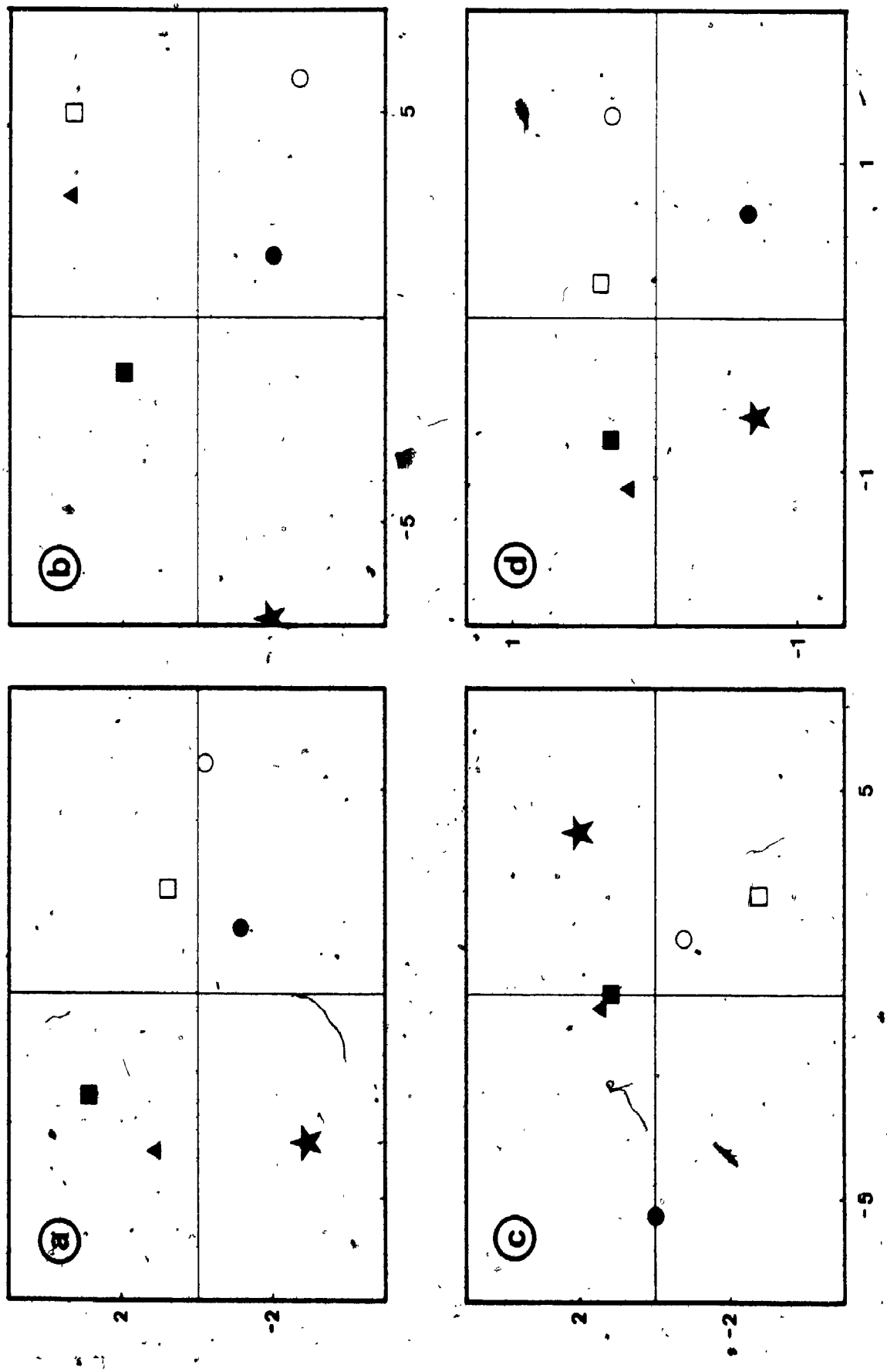

a
b
c
d
2
-2
5
-5
1
-1

Figure 8.6 Evaluation of the results of canonical variates analysis. Positions of the centroids of six C. crus-galli topodeme samples (T1 - T6), plus that of the topodeme sample of C. punctata (T7), in the space defined by:

(a) the first two canonical variates obtained in the analysis described in Table 8.11.

(b) the first two principal coordinates obtained in the analysis described in Table 8.14a, based on generalized distances calculated pairwise among topodeme samples using the pooled covariance matrix (Table 8.13a).

(c) the first two principal coordinates obtained in the analysis described in Table 8.14b, based on generalized distances calculated pairwise among topodeme samples using the common covariance matrix (Anderson, 1958; Table 8.13b).

(d) the first two principal coordinates obtained in the analysis described in Table 8.14c, based on information radii calculated pairwise among the samples (Table 8.13c).

The scales of the axes in (a - c) are the same. See Figure 8.1 for explanation of the symbols used. See text for details.

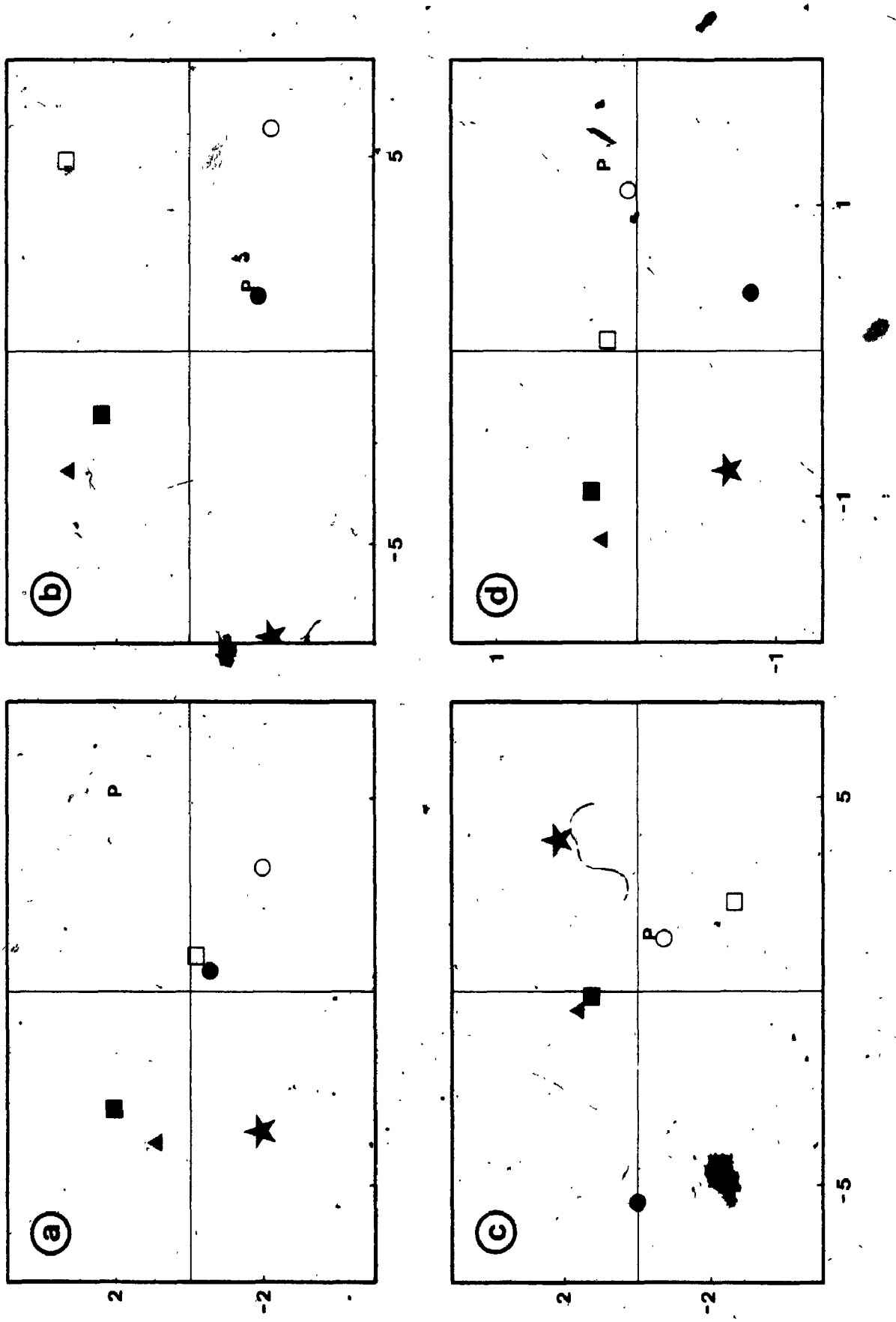

(a)
(b)
(c)
(d)
P
P
P
P
2
-2
5
-5
2
-2
1
-1

Table 8.13 Semi-matrices of the resemblance coefficients indicated, excluding the principal diagonal. Resemblances are between the centroids (mean vectors) of topodeme random samples (T1 - T7). The descriptors used in calculating resemblances were STYL, TCAL, WFL, LCAL, LFR and WFR (See Table 3.1 for explanation of abbreviations). See text for details. See Chapter 4 for details of the calculation of the resemblance coefficients. See also Figures 8.5 and 8.6 and Table 8.14.

(a) Mahalanobis' generalized distances calculated from the pooled covariance matrices for pairs of topodeme samples.

T2	6.750					
T3	6.751	3.212				
T4	5.734	8.289	10.501			
T5	6.085	6.439	7.089	5.189		
T6	13.265	6.506	5.174	13.006	8.280	
T7	4.423	6.916	6.900	4.568	3.770	7.644
	T1	T2	T3	T4	T5	T6

Table 8.13 Cont.

(b) Mahalanobis' generalized distances calculated from the common covariance matrices for pairs of topodeme samples (Anderson, 1958).

T2	3.823					
T3	4.043	0.788				
T4	1.880	2.187	2.369			
T5	7.970	3.074	2.036	6.017		
T6	4.736	1.366	1.317	3.006	9.477	
T7	2.924	2.377	3.814	1.348	5.818	2.158
	T1	T2	T3	T4	T5	T6

(c) Information radii for pairs of topodeme samples.

T2	1.160					
T3	1.266	0.353				
T4	0.998	2.061	2.377			
T5	1.078	1.743	1.950	0.915		
T6	1.370	1.033	0.816	2.134	1.119	
T7	1.431	2.332	2.564	0.743	1.442	2.230
	T1	T2	T3	T4	T5	T6

Table 8.14 Parameters of principal coordinates analyses of the resemblance matrices in Table 8.13, based on either 6 (T1 - T6) or 7 (T1 - T7) topodeme samples. In each case only four coordinate axes were found, corresponding to the eigenvalues (I - IV) shown.

(a) Mahalanobis' generalized distances calculated from the pooled covariance matrices for pairs of topodeme samples.

6 Topodeme Samples (see Figure 8.5b):

	I	II	III	IV
Eig(1)	124.67	41.34	11.56	5.36
% Tr(2)	68.15	22.60	6.32	2.93
Cum % Tr(3)		90.75	97.07	100.00

7 Topodeme Samples (see Figure 8.6b):

	I	II	III	IV
Eig(1)	127.27	43.65	15.96	8.66
% Tr(2)	65.08	22.32	8.16	4.43
Cum % Tr(3)		87.40	95.56	99.99

Table 8.14 Cont.

(b) Mahalanobis' generalized distances calculated from the common covariance matrices for pairs of topodeme samples (Anderson, 1958).

6 Topodeme Samples (see Figure 8.5c):

	I	II	III	IV
Eig(1)	52.27	15.51	1.31	0.30
% Tr(2)	75.33	22.35	1.89	0.43
Cum % Tr(3)		97.68	99.57	100.00

7 Topodeme Samples (see Figure 8.6c):

	I	II	III	IV
Eig(1)	54.19	15.70	5.09	1.27
% Tr(2)	71.07	20.58	6.67	1.67
Cum % Tr(3)		91.65	98.32	99.99

(c) Information radii for pairs of topodeme samples.

6 Topodeme Samples (see Figure 8.5d):

	I	II	III	IV
Eig(1)	4.46	1.27	0.25	0.03
% Tr(2)	74.10	21.17	4.18	0.55
Cum % Tr(3)		95.27	99.45	100.00

Table 8.14 Cont.

7 Topodeme Samples (see Figure 8.6d):

	I	II	III	IV
Eig(1)	6.23	1.35	0.77	0.20
% Tr(2)	72.81	15.77	9.05	2.37
Cum % Tr(3)		88.58	97.63	100.00

(1) Eigenvalue of the transformed distance matrix.

(2) % of the sum of the eigenvalues.

(3) Cumulative % of the sum of the eigenvalues.

8.7 Multi-group PCA

An alternative to CVA, when group dispersions are found to be unequal, is multi-group principal components analysis (Campbell, 1976; Pimentel, 1979). The seven topodeme samples (111 OTUs) were ordinated by R-PCA (Section 6.2) using for eigenanalysis the matrix W of pooled within-group dispersions output by program MANOVA (Section 7.1, 8.2; Table 8.11). Thus the transformation of the OTU vectors of (ranged) descriptor scores into vectors of component scores is accomplished by means of the eigenvectors of a dispersion matrix the variances and covariances of which do not themselves contribute to group separation. Nevertheless, the disparateness of the topodeme samples is readily apparent (Fig. 8.7), and is here a function of the original OTU descriptor scores.

Similarly, this approach to ordination of the sample also permits visualization of the contrast between the punctata OTUs (T7) and those of 10- and 20-stamen C. crus-galli (T1 - T6) with respect to the magnitude of their dispersion (Chapter 7). In view of the large contribution made to the first two components by th descriptors STYL and TCAL (Fig. 8.7; Table 8.15) it was of some concern that the contrast between T7 and the crus-galli topodeme samples seen here might be a function of only these two descriptors. This concern was especially strong with

regard to TCAL since this descriptor is (ordered) multistate (Table 3.1) and has been shown elsewhere (Table 6.1, 8.4) to have as a consequence a larger variance or coefficient of variation over the entire sample than the other five descriptors in this analysis. However, repetition of the analysis first without TCAL and then without both STYL and TCAL (see Table 7.2, 7.3, 8.2) led to results (not shown) very similar to those in Fig. 8.7, in which the scatter of the punctata topodeme sample in the plane defined by the first two principal components of W was appreciably greater than that of any one of the crus-galli topodeme samples.

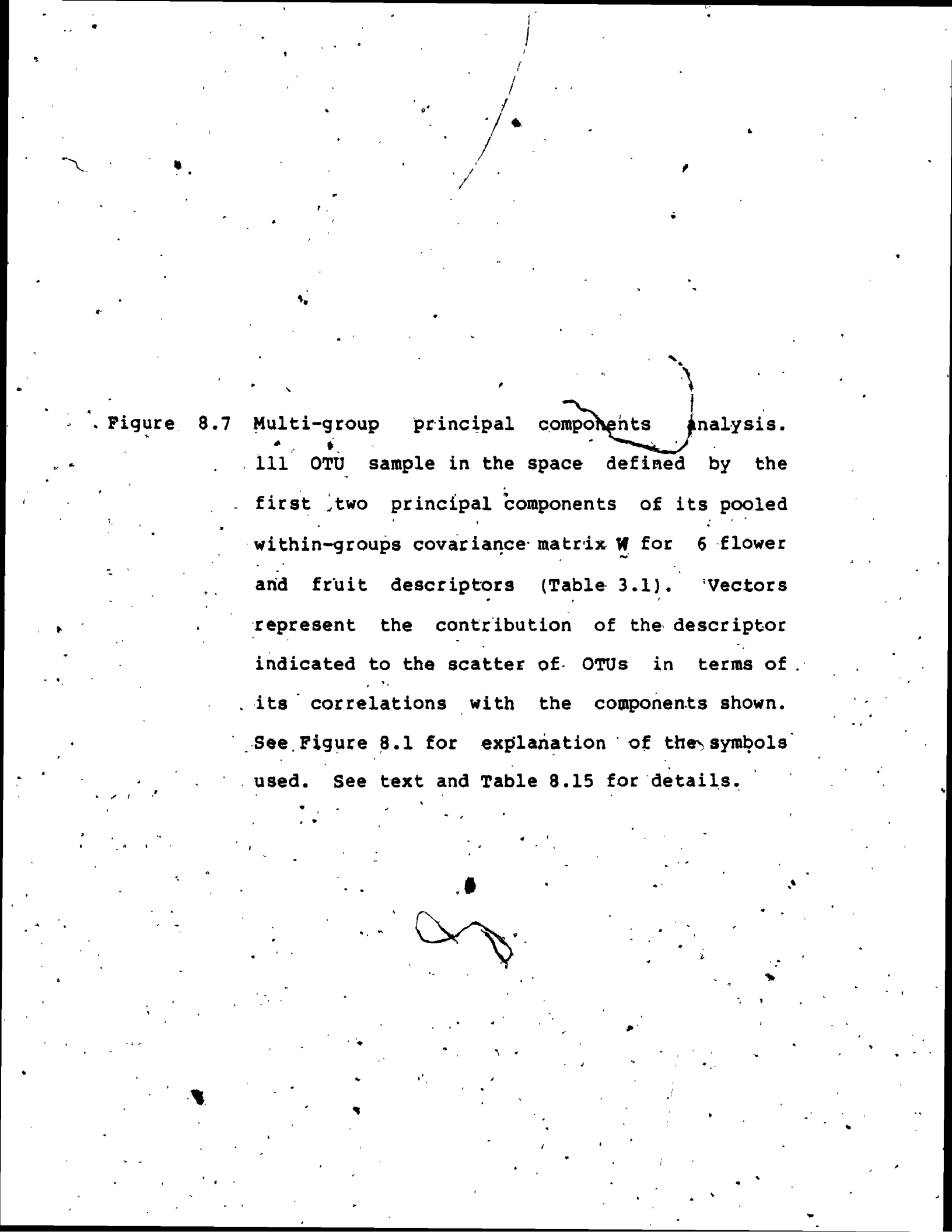


Figure 8.7 Multi-group principal components analysis. 111 OTU sample in the space defined by the first two principal components of its pooled within-groups covariance matrix $\underset{\sim}{W}$ for 6 flower and fruit descriptors (Table 3.1). Vectors represent the contribution of the descriptor indicated to the scatter of OTUs in terms of its correlations with the components shown. See Figure 8.1 for explanation of the symbols used. See text and Table 8.15 for details.

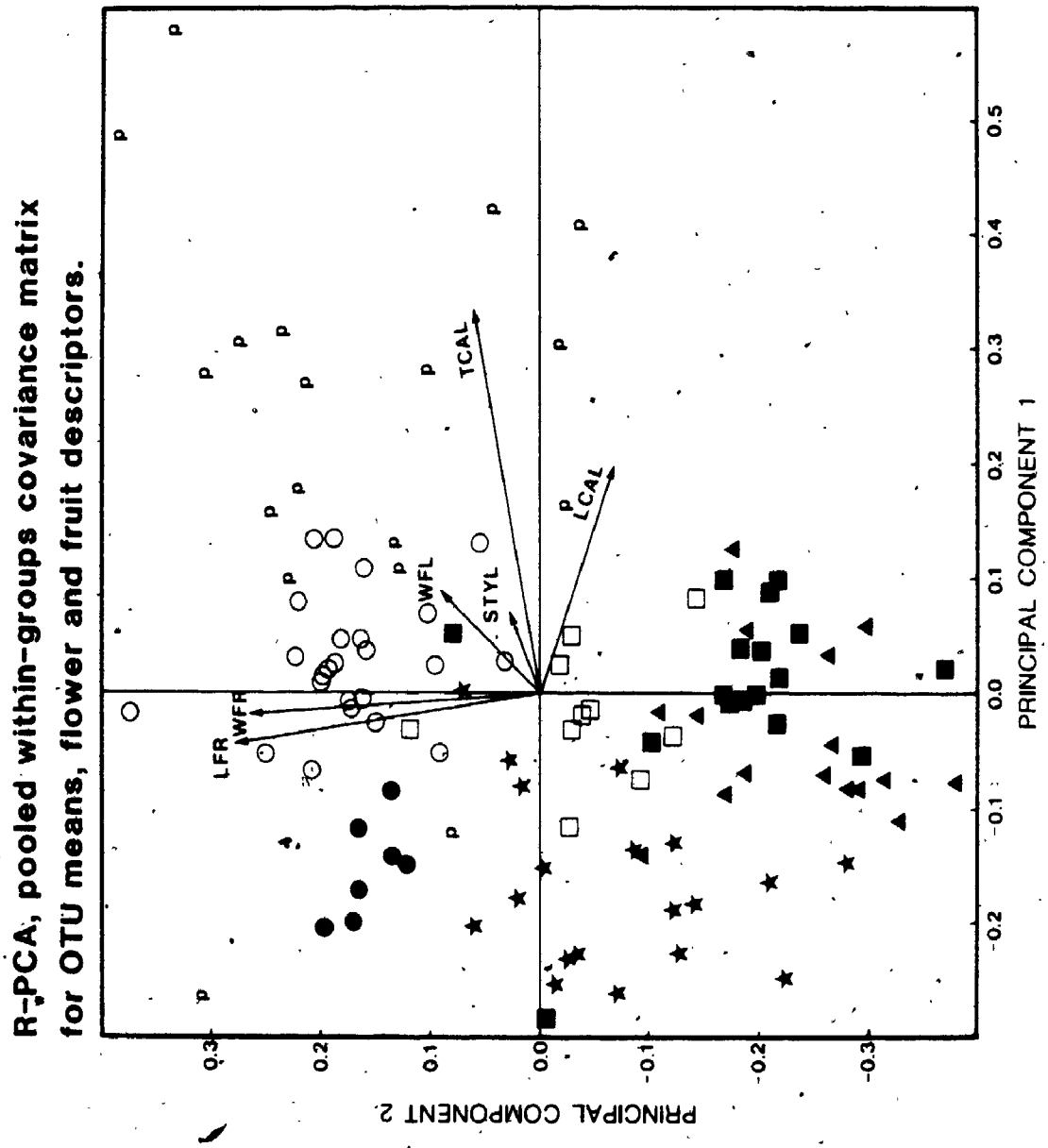

R-PCA, pooled within-groups covariance matrix
for OTU means, flower and fruit descriptors.
PRINCIPAL COMPONENT 2
PRINCIPAL COMPONENT 1
TCAL
LCAL
WFL
STYL
WFR
LFR
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2

Table 8.15 Multi-group principal components analysis of the pooled within-groups covariance matrix W for the 111 OTU sample and 6 ranged flower and fruit descriptors (Table 3.1 and Table 8.11). See text for details. See also Figure 8.7. Groups are the seven topodeme samples (T1 - T7).

	Components					
	I	II	III	IV	V	VI
Eig(1)	0.011	0.009	0.006	0.003	0.002	0.001
% Tr(2)	34.43	27.29	19.17	9.22	6.02	3.88
Cum % Tr(3)		61.72	80.89	90.11	96.13	100.00

Eigenvectors:

	I	II	III	IV	V	VI
STYL	0.167	0.067	0.032	0.186	0.747	-0.611
TCAL	0.821	0.153	-0.530	0.002	-0.142	0.041
WFL	0.214	0.227	0.476	0.770	-0.292	-0.015
LCAL	0.488	-0.165	0.694	-0.488	0.072	0.091
LFR	-0.110	0.686	0.085	-0.366	-0.357	-0.498
WFR	-0.044	0.651	0.045	-0.021	0.451	0.607

(1) Eigenvalues; those corresponding to significant components (according to Bartlett's sphericity test, Pimentel, 1979) are underlined.

(2) % trace of W.

(3) Cumulative % trace.

8.8 Leaf Spectra

Melville (1953) coined the term "leaf spectrum" to describe the characteristic assemblage of sizes and shapes exhibited by successive leaves along a stem, within a single growing season. The term thus describes the results of heteroblastic leaf development (Goebel, 1900; Greyson, Walden, & Smith, 1982). Crataegus species exhibit such leaf development, notably on their short shoots, as in C. oxyacantha L. (cf. C. laevigata (Poir.) DC.) and C. monogyna (Dau, 1941, Fig. 3, 4; Synnott, 1978) as well as in C. crus-galli s. l. and C. punctata (Fig. 8.8a-d, 8.9a-d).

Although several other methods have been suggested (e.g. Ashby, 1950; Melville, 1953; Meltzer, Searle & Brown, 1967; Więckowska, 1970), the sequential changes in leaf shape seen on short shoots of C. crus-galli and C. punctata may be represented by regressions of leaf dimensions (e.g. X, Y, Z, Table 3.3 and Fig. 3.2) or their log-transformed ratios on leaf position. This has been done for the random subsamples of OTUs from each topodeme sample, as well as for the OTUs of C. sp. aff. C. bushii, each OTU being represented by a single leaf spectrum.

The dependent variable selected was the log-transformed ratio of the leaf length above the widest point (X) to that

below the widest point (Z). This represents a contrast between obovateness (logX - logZ small) and lanceolateness (logX - logZ large). Another possibility was to use the log-transformed ratio of one or the other of the two leaf lengths (X, Z) or their sum, to the leaf width (Y). This was rejected, however, because its relationship with leaf position was markedly non-linear. Simple regression of (logX - logZ) on leaf position and pairwise comparison of the resulting slopes (Sokal & Rohlf, 1969) indicates that almost every topodeme (sub-) sample is markedly distinct from every other one (Table 6.2). Only the slopes for T6 and C. sp. aff. C. bushii are not significantly different.

The heterogeneity of slopes may also be demonstrated as an interaction between leaf position (P) and topodeme (T) in a multiple regression analysis of coariance (Table 6.3; Kim & Kohout, 1975). Here, (logX - logZ) is regressed on P, together with a series of binary-valued dummy variables ($T_1 \ldots T_6$) representing the topodeme affiliation of each leaf spectrum, and their interactions ($PT_1 \ldots PT_6$). This amounts to statistical control over variation in leaf shape with position, in comparisons between topodemes. The individuality of the topodeme slopes is indicated by the significant interaction term in the analysis of covariance table (Table 6.3).

Thus, short shoot leaf heteroblasty provides further evidence of the differentiation of the 10- and 20-stamen C. crus-galli topodeme samples from one another, as well as from the bushii and punctata samples available. The heteroblasty documented here is also the reason for the choice earlier (Section 3.5) to record the complete suite of six leaf descriptors from only a single leaf position. The terminal leaf was chosen here, rather than a specific position, in order to simplify the process of choosing which leaf to score. The first subterminal leaf has also been used, by Gostyńska-Jakuszewska (1975).

Figure 8.8 Leaf spectra of Crataegus section Crus-galli. Plot of lanceolateness (logX - logZ; see Table 3.3 for explanation of dimensions X and Z) against leaf position, from the most basal (position = 1) to terminal, for individual short shoots each representing one OTU in the 58 OTU subsample, and belonging to topodeme samples as follows:

(a) 20-stamen C. crus-galli, Site 1 (T1);

(b) 10-stamen C. crus-galli, Site 1 (T2);

(c) 20-stamen C. crus-galli, Site 2 (T4);

(d) 10-stamen C. crus-galli, Site 2 (T5);

Regression line (Table 8.16) is solid line; dashed lines indicate corresponding 95% confidence region. A typical leaf spectrum is illustrated in each case; the scale bar at the right represents 1.0 cm.

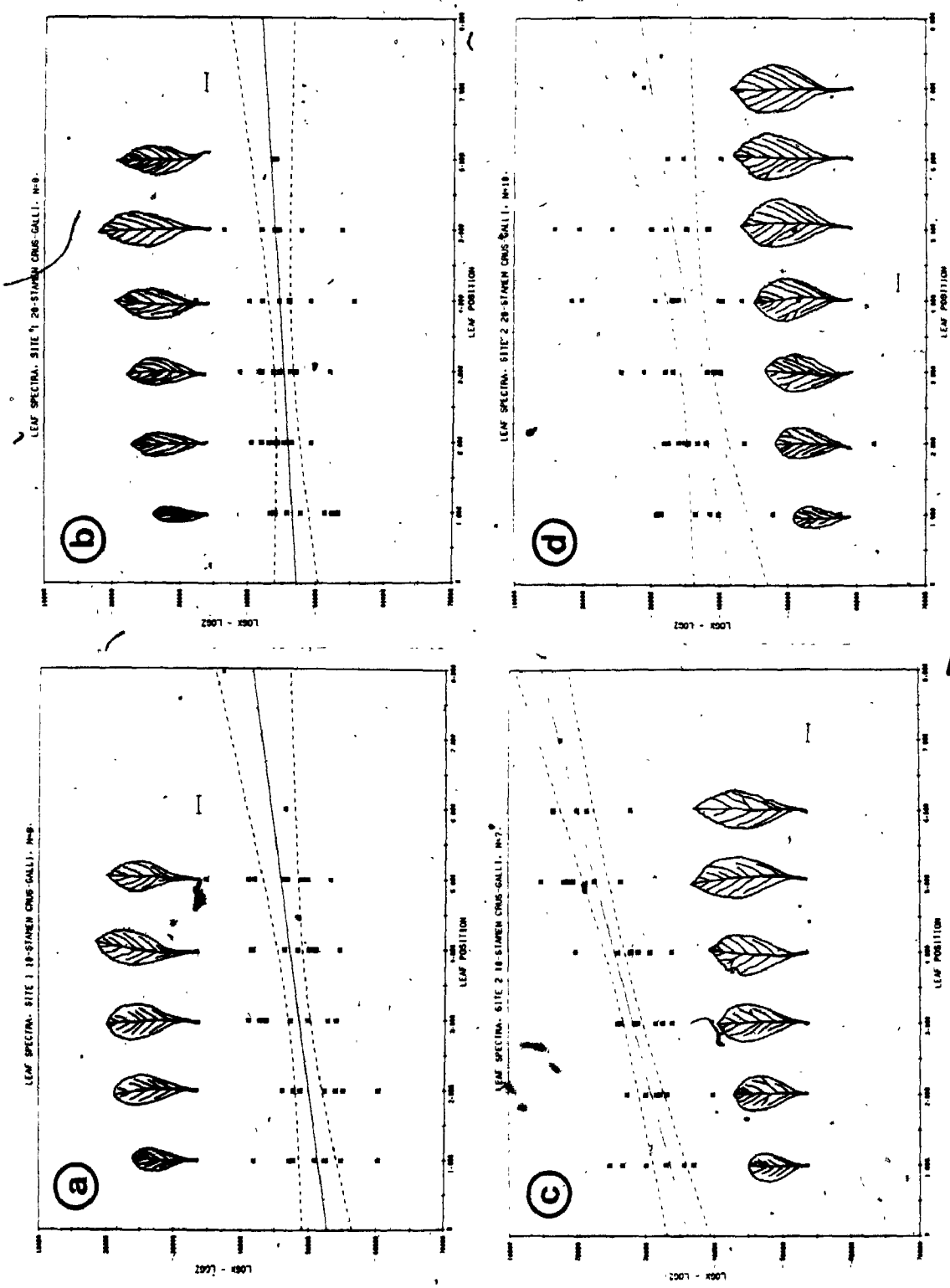

LEAF SPECTRA- SITE 1 20-STAMEN CRUS-GALLI. N=9.
LEAF POSITION
LOGX - LOG2
a
LEAF SPECTRA- SITE 2 20-STAMEN CRUS-GALLI. N=10.
LEAF POSITION
LOGX - LOG2
b
LEAF SPECTRA- SITE 1 10-STAMEN CRUS-GALLI. N=8.
LEAF POSITION
LOGX - LOG2
c
LEAF SPECTRA- SITE 2 10-STAMEN CRUS-GALLI. N=7.
LEAF POSITION
LOGX - LOG2
d

Figure 8.9 Leaf spectra of Crataegus section Crus-galli and section Punctatae. Plot of lanceolateness (logX - logZ; see Table 3.3 for explanation of dimensions X and Z) against leaf position, from the most basal (position = 1) to terminal, for individual short shoots each representing one OTU in the 58 OTU subsample, and belonging to topodeme samples as follows:
(a) 10-stamen C. crus-galli, Site 3 (T3);
(b) 10-stamen C. crus-galli, Site 4 (T6);
and (d) C. punctata, Site 5 (T7).
In (c), data for deterministically sampled OTUs C. sp. aff. C. bushii (Site 6). Regression line (Table 8.16) is solid line; dashed lines indicate corresponding 95% confidence region A typical leaf spectrum is illustrated in each case; the scale bar at the right represents 1.0 cm.

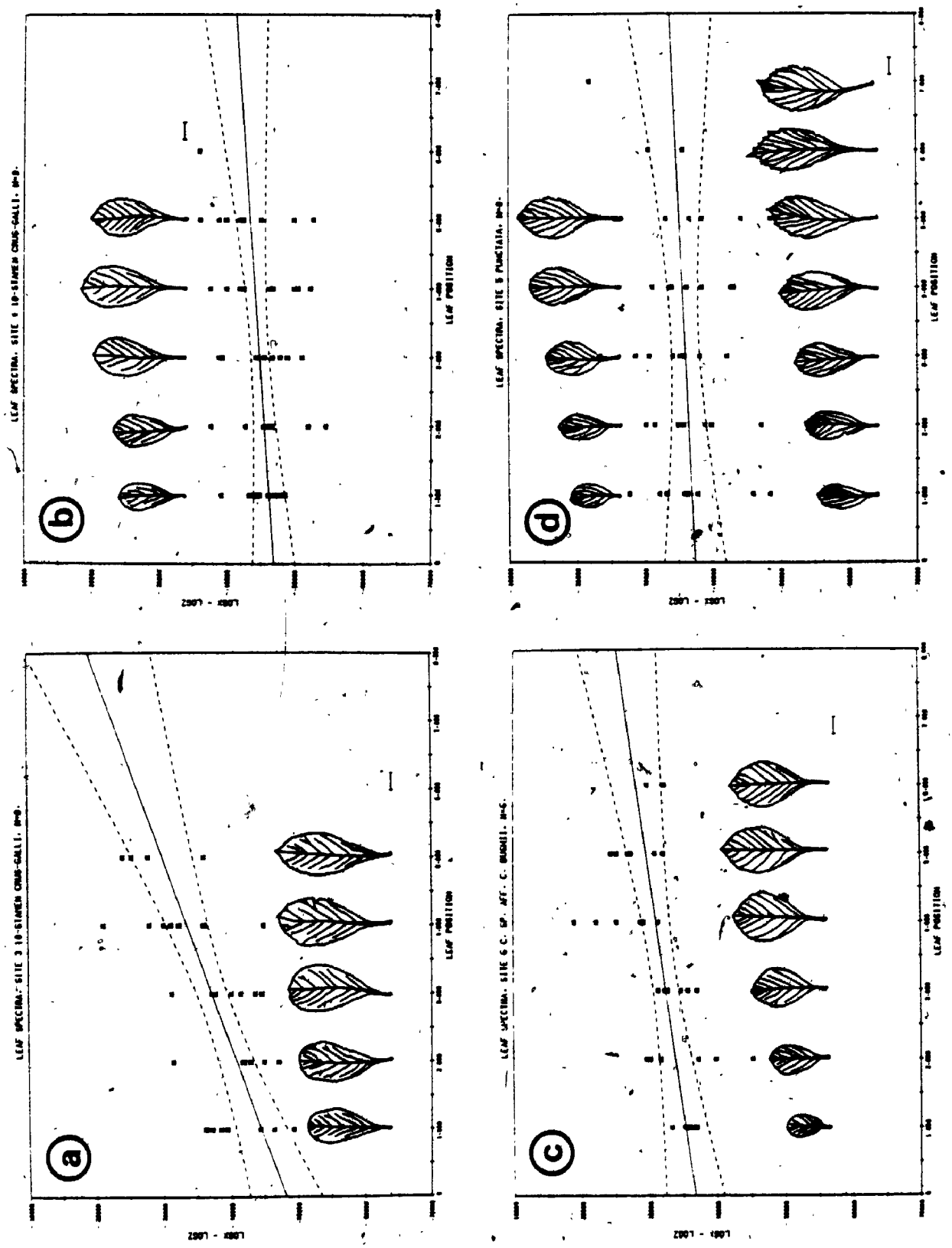

a
b
c
d
LEAF POSITION

Table 8.16 Comparison of topodeme slopes for regression of leaf shape described by (logX - logℓ) on leaf position. Entries are F-ratios for the pairwise comparison of regression coefficients (Sokal & Rohlf, 1969). The sample of C. sp. aff. C. bushii (Fig. 8.9c) is included for comparative purposes.

T2	151.3						
T3	354.3	235.7					
T4	113.2	28.4	131.1				
T5	548.3	292.7	122.6	121.9			
T6	9.3	144.2	356.6	112.4	537.0		
T7	37.6	140.4	291.6	128.4	412.4	44.9	
bushii	145.1	3.2	213.2	23.9	268.7	138.8	130.7
	T1	T2	T3	T4	T5	T6	T7

All slopes significantly different ($p < 0.001$; for T1 and T6, $0.05 > p > 0.001$) except for T2 and bushii ($p > 0.05$).

Regression Statistics:

	T1	T2	T3	T4	T5	T6	T7	bushii
n(1)	42	41	36	54	40	46	43	33
a(2)	-0.47	-0.53	-0.48	-0.42	-0.36	-0.47	-0.37	-0.37
b(3)	0.674	1.412	3.645	1.696	2.693	0.713	0.448	1.432
SSreg	0.004	0.018	0.083	0.040	0.084	0.005	0.002	0.017
SSres	0.080	0.098	0.155	0.396	0.071	0.090	0.178	0.087
b(4)	ns	*	***	*	***	ns	ns	*

Table 8.16 Cont.

(1) Number of leaves per topodeme sample.

(2) Intercept.

(3) Slope x 100.

(4) Significance of b.

ns not significant ($p > 0.05$)

* $0.05 > p > 0.01$

*** $p < 0.001$

Table 8.1 Analysis of the covariance of (logX - logZ) with leaf position (P) in seven topodeme samples. The influence of topodeme (T) and its interaction with position (PT) is incorporated by means of binary-valued dummy variables, as described in the text.

Source	SS	d.f.	MS	F
Regression of (logX-logZ) on P, T and PT:	1.536	13	0.118	30.458***
...on P and T alone:	1.485	7	0.212	54.714***
...on their interaction (PT) alone:	0.050	6	0.008	2.159*
Residual:	1.117	288	0.004	
Total:	2.653	301		

* $0.05 > p > 0.01$

*** $p < 0.001$

8.9 Discussion

The total sample of 160 OTUs is differentiated into groups not only with respect to the 11 flower and fruit descriptors (Chapter 5) but also (the 60 OTU subsample) with respect to the leaf descriptors (Chapter 6). With regard to the 111 randomly sampled 10- and 20-stamen crus-galli plus punctata OTUs these groups correspond largely to the seven topodeme samples (T1 - T7). In Chapter 7 the covariance matrices of these samples were compared and found to be similar in the case of the leaf descriptors, and to differ in that of the flower and fruit ones. The differences found were principally in the magnitude of the covariance matrix determinants, and only to a lesser degree in the pattern of descriptor covariation. In the present chapter canonical analyses have been used to depict and evaluate the distinctness of the topodeme samples (Section 8.2, 8.4, 8.5). Ordination methods (PCoA, PCA) were used to validate the canonical analyses (Section 8.6, 8.7). Finally, linear regression techniques applied to the leaf data available for the 60 OTU subsample also confirmed the distinctness of the topodeme samples (Section 8.8).

What follows is mainly a discussion of the canonical analysis results. As will be pointed out in the concluding part of this section, the data that have been analyzed here fail to satisfy at least some of the assumptions made when

canonical analysis is used in formal hypothesis testing. However, in addition to this usage canonical analysis is also a valuable tool for data description and hypothesis formulation (Campbell, 1978; Gittins, 1979; Pimentel, 1981). In this role many of its requirements of the data may be relaxed, notably that of multivariate normalilty (Campbell, 1978).

The random sample available here is limited in size and is marked by extreme heterogeneity of the topodeme sample covariance matrices. This is a further reason why the results obtained from canonical analyses that relate to the differentiation of the topodeme samples are treated below as being merely descriptive of the sample and thus having only limited predictive power concerning the populations from which the samples were drawn.

8.9.1 Multivariate Analysis of Variance

In Section 8.2 it was seen how multivariate analysis of variance of both the flower and fruit and the leaf data leads to rejection of the null hypothesis of the equality of topodeme sample mean vectors (Table 8.1, 8.2). Although an idea of the relative contribution of individual descriptors to group discrimination may be obtained from their univariate F-ratios (Table 8.3) ordinations give more insight into the relationships between groups and

descriptors.

8.9.2 Correlation of the Leaf and Flower Data

In Chapter 6 the two data sets available for the 60 OTU subsample presented a sharp contrast, with the flower and fruit data showing much stronger topodeme sample differentiation than the leaf data. This contrast appears to result from function of the larger topodeme sample dispersions for the leaf descriptors (Chapter 7). Nevertheless it was found in Section 8.4 that the two data sets are highly correlated (Table 8.5). Moreover, the nature of this correlation was such that plotting the OTUs in the spaces defined by the first two canonical variates of both domains resulted in fairly distinct clusters of OTUs belonging to the same topodeme sample (Fig. 8.1). Thus the space in which the two data sets are most highly correlated (defined by the eigenvectors of $\underset{\sim}{R}_{xx}^{-1}\underset{\sim}{R}_{xy}\underset{\sim}{R}_{yy}^{-1}\underset{\sim}{R}_{yx}$) is one in which the differences among topodeme samples in their descriptor scores result in a considerable degree of group discrimination. Variation in both sets of descriptors is highly correlated with taxon and site.

8.9.3 Topodeme Sample Differentiation

Canonical variates analyses were used to study the way in which each of the data sets available for the 60 OTU

subsample distinguished the topodeme samples (Section 8.5). In both cases topodeme samples were readily distinguished (e.g. Fig. 8.4). As in the CCA discussed above (Fig. 8.1, 8.2; Table 8.6, 8.7), however, the six leaf descriptors contributed nearly equally to group discrimination (Fig. 8.3; Table 8.8d). Contrasts in leaf shape between the two crus-galli morphotypes were evident in both analyses. Both the CCA and the CVA of the flower and fruit data are strongly polarized by the presence of the punctata topodeme sample (T7), however, and this appears to be responsible for the unequal importance of the descriptors. Group discrimination in the plane of the first two CVA canonical variates (Table 8.10d) and those of the CCA flower-fruit domain (Fig. 8.1;Table 8.6, 8.7) is mainly a function of descriptors STYL and WFL (also LFR in the CCA). The contribution of the flower and fruit descriptors to topodeme sample discrimination is illustrated more clearly in the larger sample for which this data is available, however.

8.9.3.1 Descriptor Contributions to Group Discrimination

With the entire random sample of 93 crus-galli OTUs the contrast between the 10- and 20-stamen morphotypes seen in the plane of the first two canonical variates (Fig. 8.5a) is a function of style number and other correlates of stamen

number, such as flower and fruit width, as well as fruit length (Section 8.5.2.1; Table 8.9d). While the contribution of calyx lobe toothing in this plane is appreciable too, it is almost normal to the trend between the two morphotypes. In this analysis the contribution of almost all the descriptors to the first two canonical variates results in only minimal overlap of the OTUs of each topodeme sample in this plane (not shown).

Addition of the punctata topdeme sample (T7) polarizes the analysis (Section 8.5.2.2). The first canonical variate is dominated by a crus-galli-punctata contrast in style number (Fig. 8.4). Differentiation of the crus-galli morphotypes is along both the first and second canonical variates, associated with a further contrast in flower, and to a lesser extent, fruit width (Fig. 8.4; Table 8.11e). Discrimination of the individual crus-galli topodeme samples appears to involve all descriptors except calyx lobe length. However, because of the presence of the punctata sample, this discrimination is complete only in the space of the first three canonical variates.

Multi-group PCA based on the pooled within-groups covariance matrix W (Section 8.7) demonstrates that even in a space dominated by the descriptors whose variation does not discriminate between the topodeme samples, these samples nevertheless remain largely distinct (Fig. 8.7).

Differentiation between crus-galli and punctata here appears in general to be a function of the longer and more toothed calyx lobes of the latter species (Fig. 8.7). Differentiation of the crus-galli morphotypes is much less evident than in the CVA.

8.9.3.2 Eigenvector Comparisons

PCA of the pooled within-groups covariance matrix discussed above, when performed on the 93 OTU crus-galli random sample (not shown) provides six reference vectors (eigenvectors of $\underset{\sim}{W}$) characteristic of the crus-galli sample as a whole that may be used in comparisons with each of the topodeme samples. The basis for comparison is Anderson's test of the equality of an observed eigenvector and a reference vector (Anderson, 1963; Seal, 1964):

$$X^2 = N(\lambda_i \underset{\sim}{u}' \underset{\sim}{S}^{-2} \underset{\sim}{u} + \lambda_i^{-1} \underset{\sim}{u}' \underset{\sim}{S}^{2} \underset{\sim}{u} - 2),$$

where N is the sample size, $\underset{\sim}{S}$ the sample covariance matrix with eigenvalues λ_i, i = 1...t components, and $\underset{\sim}{u}$ is the reference vector. The null hypothesis is that the reference vector could be an eigenvector of $\underset{\sim}{S}$ is evaluated by comparing X^2 with χ^2 for t - 1 degrees of freedom.

The eigenvectors of $\underset{\sim}{W}$ for the 93 OTU sample and the results of Anderson's test for each topodeme sample are given in Table 8.18. As in Chapter 7, results for T5 (N = 7) are discounted because of its excessively small

sample size (Seal, 1964). For the remainder of the topodeme samples Table 8.18b suggests a differentiation first of the punctata sample (T7) and then of T6 from the remaining crus-galli samples. The results for T7 are expected in view of the source of the reference vectors. The differentiation of T6 recalls the way in which this sample was distinguished in M-S and residual plots of arctanh correlations in Chapter 7, as well as in some combinations of resemblance function and sorting algorithm in Chapter 5 (Fig. 5.1, 5.2) and Chapter 6 (Section 6.3).

8.9.4 Limitations of Canonical Analyses

Successful use of canonical analysis requires attention to a number of factors that are discussed by Gittins in his monograph on the method (Gittins, 1979). Descriptors must be chosen that are in fact relevant to the phenomena of interest. As an optimal ratio of descriptors to sample size must be achieved to avoid overfitting the model (see below) nothing is gained by including all possible descriptors. Moreover, descriptors highly correlated with each other do not add to explanatory power and may increase the sensitivity of the analysis to sampling error (Gittins, 1979).

The applicability of the canonical analysis methodology used here is generally limited to data structures that are

continuous and linear. Further, statistical inference from canonical analyses requires that departures from multivariate normality, as well as from homogeneity of group covariance matrices, be as small as possible (Gittins, 1979). Extreme outliers may also vitiate analyses.

Guidelines proposed by Barcikowski and Stevens (1975) and Thorndike (1978) suggest that a ratio of descriptors to sample size in the range 0.05 - 0.07 is likely to enable success in the predictive use of canonical analysis (Gittins, 1979). Results obtained with larger ratios are likely instead to be descriptive only of the sample from which they came. In such cases the degree of relationship between data sets, described by their canonical correlations, is likely to be exaggerated (Gittins, 1979). Canonical weights are especially unstable and sensitive to the effects of overfitting and sampling error.

In the present study some care has been taken in selecting the best suited descriptors for the analyses described here (Chapter 7). Exploratory studies of the data (Section 3.7; Chapter 6) suggest that the descriptors used do in fact approximate the requirements of continuity and linearity. Similarly, extreme outliers do not appear to be present. However, in the total random sample of 111 OTUs the descriptors studied here (six flower and fruit, and six leaf descriptors) do show some departures from univariate

normality (Section 3.7). As documented in Chapter 7 as well as in this one (Fig. 8.4, 8.7) it is evident that the topodeme samples are not homogeneous with respect to their covariance structures.

It has also become apparent that the available samples are not all as large as they should be. The behavior of the arctanh correlations for T5 makes this clear (Section 7.5). Of the analyses discussed so far, only that of the 155 OTU sample (Section 8.5.2.3; Table 8.12) in which random and non-random samples have been pooled to form taxonomic groups leads to a ratio of descriptors to sample size within the recommended range ($d/s = 0.065$; canonical correlation = 0.92; Wilks' $\Lambda = 0.0088$). In the other analyses described the minimum value of this ratio is 0.11 (for the 111 OTU sample). The very high values obtained for canonical correlations in these other analyses (0.96 - 0.98) and the correspondingly small values of Wilks' Λ (0.0010 - 0.0035) suggest that the degree of differentiation of the topodeme samples seen is probably overestimated to some degree.

At the same time the results of PCoA ordination of the topodeme sample centroids on the basis of the generalized distances and information radii between them (Section 8.6) are encouraging. They indicate that although covariance matrix heterogenity and disparity in sample size have an

effect on the spatial relationships among the topodeme sample centroids in the PCoA space, the separation of the samples that is found is comparable to that seen in the corresponding CVA (Fig. 8.5, 8.6).

In addition, CVA results were most highly correlated with the PCoA of information radii. The calculation of information radii is independent of assumptions concerning distributional and covariance properties of the data. Accordingly, the high correlation found in Section 8.6 suggests that despite the violation of assumptions concerning homogeneity of group covariance structures and multivariate normality of the data, the ordination of group centroids produced by CVA is a realistic one.

The extreme separation of group centroids that may occur in overfitted models (Oxnard, 1978) does not appear to be present in the analyses presented here, where the maximum separation found is on the order of 5 - 6 standard deviation units over a total range of 10 - 15 such units (Fig. 8.4).

Likewise, the structure coefficients and descriptor communalities have provided an alternative to the canonical weights for evaluation of the contributions of the individual descriptors to topodeme sample differentiation in the various analyses. The pictures of the analyses given by the canonical weights (Fig. 8.1, 8.3, 8.4) turn out to be

largely corroborated by these other, more stable interpretive measures (compare respectively Fig. 8.2 and Table 8.8, 8.11).

Finally, as a check on the extent to which the results obtained here are discounted by reason of the limitations (especially of sample size) discussed above, canonical analysis of the 111 OTU sample was performed using a simplified taxonomic grouping hypothesis (10-stamen C. crus-galli, N = 59; 20-stamen C. crus-galli, N = 34; C. punctata N = 18). Both $H_0: \underset{\sim}{\Delta}_k = \underset{\sim}{\Delta}$ and $H_0: \underset{\sim}{\mu}_k = \underset{\sim}{\mu}$ for k = 1...3 were rejected ($p < 0.001$). The two canonical roots which were obtained (corresponding to the eigenvalues of $\underset{\sim}{W}^{-1}\underset{\sim}{A}$) were both significant according to Bartlett's test and the union-intersection test ($p < 0.001$). The configuration of OTUs in the plane of the two canonical variates resembled that in Fig. 8.4 not only in the relative positions of the three taxa but also in the disposition of the individual topodeme samples. The principal difference found was in the way T1 and T4, and also T2 and T3 were not well differentiated from each other. Separation of the punctata sample from that of the crus-galli morphotypes occurred mainly along the first canonical variate. Separation of the crus-galli morphotypes took place along both canonical variates. Descriptor communalities for these canonical variates were highest for

style number (H = 0.964, due mainly to the first) and flower width (H = 0.867, due to the first, and to a lesser extent also to the second canonical variate). Thus apart from failure to resolve two pairs of topodeme samples, this analysis in which the ratio of descriptors to sample size was 0.072 repeats all the main features of those described in detail in this chapter which were concerned with topdeme sample differentiation.

8.9.5 Conclusions

The studies reported in this chapter have corroborated the breakdown of the random sample into taxonomic groups that was suggested earlier (Chapter 5, 6). The groups formed in this way are the two common morphotypes of C. crus-galli L. sensu-lato (10-stamen, 20-stamen) and C. punctata Jacq.

In addition, the following points can also be made. Alternative hierarchical relationships were seen in the cluster analyses in Chapter 5 and in Chapter 6. Depending on the data used, and on resemblance function and sorting algorithm, 20-stamen C. crus-galli first joined either to 10-stamen C. crus-galli (pattern A) or to C. punctata (pattern B). These relationships expressed themselves in the canonical analyses in the intermediacy of the topodeme samples of the 20-stamen morphotype, between those of the

10-stamen one and the punctata sample. This intermediacy was exhibited without reference to stamen number and even in the analyses based on leaf descriptors. The significance of this observation is limited, however, by the small numbers of punctata OTUs in the comparison.

Topodeme samples are distinct from one another, although not to the same degree as the taxonomic groups described above. The crus-galli sample from the Niagara peninsula (T6) is exceptional in this regard, since it is differentiated from the other 10-stamen OTUs in all analyses. Similar material from this site was distinguished from other Ontario C. crus-galli as C. crus-galli var. pyracanthifolia Ait. by Sargent (1908). In general, however, the small sample size limits the confidence with which the crus-galli topodemes studied here can be said to be markedly differentiated from each other. Nevertheless, the indications of differentiation that are given in all the analyses are of considerable interest in view of the way in which they suggest the possible basis for the narrow species concepts applied in earlier North American treatments of the genus. In particular they bear out the observations of Rickett who distinguished forms comparable to those encountered here among both Missouri C. section Pruinosae and C. crus-galli s. l. (Rickett, 1936, 1937). Furthermore, the evidence of local differentiation is

especially valuable for its bearing on questions concerning reproductive behavior, dispersal and topodeme establishment. None of these processes are at present well understood in Crataegus, but explanations of each one, for C. crus-galli s. l. in Ontario, must be made consonant with the patterns of variation and variability that have been observed here.

Table 8.18 Comparison of the eigenvectors of the pooled within-groups covariance matrix $\underset{\sim}{W}$ calculated for the random sample of 93 OTUs of C. crus-galli s. l. with those of the covariance matrices of the individual topodeme samples (T1 - T6) making up this sample, as well as with those of the covariance matrix of the C. punctata topodeme sample (T7). The test criterion is that due to Anderson (1963). See text for details. See Table 3.1 for explanation of the abbreviations of the descriptors used.

(a) Pooled within-groups covariance matrix $\underset{\sim}{W}$ for 93 OTUs. Eigenvalues significantly different from those to the right are underlined (see below for test criterion).

	Components					
	I	II	III	IV	V	VI
Eig(1)	0.007	0.006	0.003	0.002	0.001	0.001
% Tr(2)	33.09	29.79	16.32	11.43	5.28	4.10
Eigenvectors:						
STYL	0.065	0.079	-0.070	-0.204	0.626	0.742
TCAL	0.021	0.234	0.960	-0.152	0.004	0.019
WFL	0.241	0.640	-0.048	0.721	0.094	0.024
LCAL	-0.059	0.712	-0.264	-0.596	-0.250	-0.048
LFR	0.723	-0.147	0.006	-0.054	-0.543	0.396
WFR	0.640	-0.025	-0.037	-0.240	0.492	-0.538

Table 8.18 Cont.

(b) Results of the comparison of the eigenvectors in (a) with the corresponding topodeme sample covariance matrix eigenvectors. Entries are the values of X^2 obtained for each comparison, using Anderson's test criterion. Entries corresponding to significant components (according to Bartlett's sphericity test; Pimentel, 1979) are in boldface.

	Components					
	I	II	III	IV	V	VI
T1	3.6ns	19.2***	5.0ns	31.6***	9.2ns	14.3**
T2	22.7***	13.7*	10.0ns	12.6*	6.0ns	12.3*
T3	11.7*	8.5ns	8.5ns	9.1ns	9.2ns	10.8ns
T4	17.0***	15.6**	16.5**	3.2ns	3.5ns	5.3ns
T5	772.9***	37.2***	19.3***	27.7***	186.7***	282.4***
T6	16.2**	28.2***	30.4***	42.1***	13.0*	12.6*
T7	24.9***	94.3***	28.4***	111.2***	33.5***	72.2***

ns not significant ($p > 0.05$)

* $0.05 > p > 0.01$

** $0.01 > p > 0.001$

*** $p < 0.001$

(1) Eigenvalue

(2) % Trace

CHAPTER NINE

VARIATION IN CHROMOSOME NUMBER

9.1 Introduction

As pointed out in Section 1.9 not only the genus Crataegus as a whole but also individual species complexes within it such as C. crus-galli L. sensu lato exhibit variation in ploidy level (Muniyamma & Phipps, 1979b). Crataegus crus-galli s. l. was chosen for study here in part because it had been identified by Muniyamma and Phipps as a complex of diploid, triploid, and tetraploid individuals. In the present study chromosome numbers have been determined principally for individuals for which data on reproductive behavior have been obtained (Chapter 10), in order to provide a basis for the interpretation of that data. In this way the cytological data obtained also represent a subsample of each of the strata in the total sample of 160 OTUs for which morphometric data have been analyzed in Chapters 5-8.

The study described here is limited to a description of the variation observed in the sample with respect to two aspects of the nucleotype (Bennett, 1971), namely chromosome number and size. In general, Crataegus resembles other

Maloideae in which polyploidy occurs in having many (2n = 34, 51, 68) small (< 5 µm long) chromosomes (Moffett, 1931; Byatt & Murray, 1977; Muniyamma & Phipps, 1979b).

In view of the small size and relatively large numbers of chromosomes found in Crataegus emphasis has been placed here on obtaining chromosome counts from somatic divisions. Logistical considerations also entered into the decision to limit the scope of this part of the study. The timing of meiosis in the taxa studied here was not known, initially. Muniyamma and Phipps (1979b) did not report details of the developmental stages from which they obtained meiotic stages. However, they did establish that meiosis in Ontario Crataegus taxa is brief and more or less synchronous within an individual and within topodemes. By repeated collecting it was possible to bracket meiosis at a few sites, and so characterize its timing morphologically and in terms of temperature sum (Appendix 7). However, in general it appeared that actively dividing somatic cells from the shoot apices and leaf primordia of expanding buds, and from sterile floral structures were likely to be available for longer periods than meiotic cells.

9.2 Materials and Methods

Mitotic metaphases were obtained principally from meristematic tissue at the center of vegetative shoot buds shortly after they began to open. In some cases mitoses were also obtained either from individual flower buds of unexpanded inflorescences or from pre-meiotic anther tissue. Stages of meiosis were obtained from the pollen mother cells (PMCs) of flower buds. These were collected during a short period of time three to four weeks prior to anthesis.

Both meiotic and mitotic metaphases were accumulated by treatment of longitudinally bisected flower or shoot buds with either colchicine (0.33% aqu.) or p-dichlorobenzene (PDCB; saturated aqueous solution), or by incubation of cut branches (their cut ends in water) at 5° C for 2 - 12h prior to fixation (Darlington & LaCour, 1976; Sharma & Sharma, 1972). In the case of colchicine and PDCB treatment times were 1 - 3h. Material was also fixed without pretreatment.

Fixation was done in Farmer's fluid (absolute EtOH, glacial acetic acid 3:1) chilled on ice. A saturated aqueous solution of $FeCl_3$ was added dropwise until the fixative was a pale straw color; the ferric ions served to premordant the tissue for acetocarmine staining (Darlington & LaCour, 1976). After 24h material was transferred to 70% EtOH and stored in the cold (Darlington & LaCour, 1976).

Material for Feulgen staining was hydrolyzed in 5N HCl at room temperature for 1h, rinsed in water, and transferred to leucobasic fuchsin to stain for 5 - 6h (Berlyn & Miksche, 1976). Tissue was then macerated using a sequence of pectinase (fungal, technical powder, ICN Pharmaceuticals), cellulase (fungal, powder, Calbiochem- Behring Corp.), and 45% acetic acid. Both enzymes were used at 5% (w/v, aqu.). for 5h in the case of pectinase, and for 0.5h in the case of cellulase. Material was accumulated in 45% acetic acid and occasionally held overnight at this point if necessary.

With or without feulgen staining, small amounts of macerated material were stained in hot acetocarmine (Clark 1973) and spread with a flat-tipped needle. A coverslip was applied and the cells were separated by tapping with the handle of the needle. Preparations were checked under the microscope, and those which were well spread were then flattened using a mechanical press (Micro Press, E. F. Ikonen, Winnipeg, Manitoba). The edges of the coverslip were then sealed with Permount (Fisher Scientific Co.) diluted with xylene so as to flow under the edge of the coverslip and completely seal the preparation.

Meiotic or mitotic figures were located by scanning the preparation under the low power on the microscope. Suitable ones were photographed usually with both the 60x and 100x (oil) planapochromatic objectives on a Nikon Optiphot

microscope equipped with AFM automatic exposure control. The film used was Kodak Technical Pan 2415 exposed at ASA 100 and developed for 8 minutes at 21°C with Kodak HC-110 developer diluted 1:30. All counts and measurements were made from enlarged prints of these negatives.

Herbarium vouchers of the individuals for which chromosome numbers are reported in Table 9.1 have been deposited in the Herbarium of the Department of Plant Sciences, University of Western Ontario (UWO).

9.3 Results

Counts have been made from somatic material for 28 OTUs (Table 9.1). Of these, two (OTUs 131, 731) are individuals also examined by Muniyamma and Phipps (1979b). The tetraploid count for OTU 731 (Table 9.1) is at variance with the diploid one obtained by Muniyamma and Phipps (1979b). Two counts are for individuals of C. punctata (OTUs 514, 519). The remainder are new counts for Crataegus crus-galli s. l.. Two additional individuals of Crataegus crus-galli s. l. for which Muniyamma and Phipps reported chromosome numbers are also included in the 160 OTU sample (OTUs 721, 722; Table 9.1).

Feulgen staining of the chromosomes studied here was variable and usually inadequate by itself. However,

acetocarmine gave consistently satisfactory results when used alone or in addition to feulgen staining.

The mitotic metaphase chromosomes of both Crataegus crus-galli s. l. (Plate 2: Plate 3c-f) and C. punctata (Plate 3a, b) are small, in the range 1-3 µm. This is comparable to sizes reported by Moffett (1931) and Byatt and Murray (1977) for European species of Crataegus and by Muniyamma and Phipps (1979b) for Ontario material.

There is no correlation between stamen number and level of ploidy in C. crus-galli s. l. (Table 9.2). An individual of C. sp. aff. C. bushii at site 6 was determined to be tetraploid (Plate 21; Table 9.1). Two individuals of male-sterile C. ?grandis (Site 4) were found to be triploid (Plate 3c, d; Table 9.1).

Observations on meiosis were made on material of both triploid (OTUs 105, 755) and tetraploid (OTU 752) individuals of Crataegus crus-galli s. l. OTUs 105 and 752 are both 20-stamen C. crus-galli, while OTU 755 is 10-stamen C. crus-galli. Stages of diakinesis observed in OTUs 752 and 756 show what appear to be connections between chromosomes or pairs of chromosomes These are very suggestive of some kind of ordering of the chromosomes within the meiotic prophase nucleus. Metaphase figures (Plate 3j) were difficult to interpret because little

difference was seen among univalents, bivalents and multivalents in the limited amount of material examined.

Plate 2 Mitotic metaphases of tetraploid *Crataegus* taxa, all to the same scale. 10-stamen *C. crus-galli*: (a) OTU 111, colchicine pretreatment; (b) OTU 123; (c) OTU 302, colchicine pretreatment; (d) OTU 307; (e) OTU 401; (f) OTU 424; (g) OTU 606; (h) OTU 610; and (i) OTU 731. 20-stamen *C. crus-galli*: (j) OTU 714 and (k) OTU 753. *C.* sp. aff. *C. bushii*: (l) OTU 801. See text and Table 9.1 for additional details.

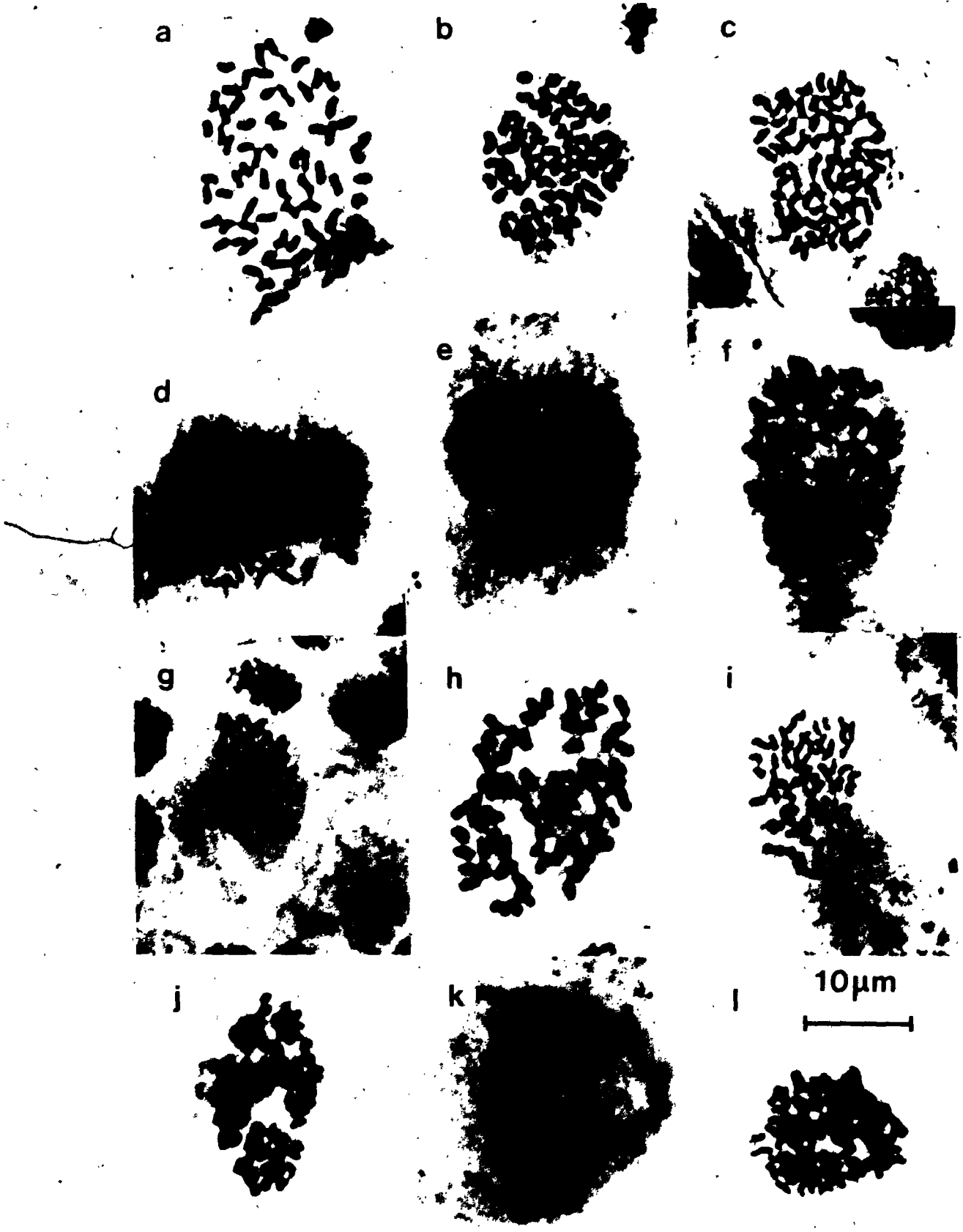

a
b
c
d
e
f
g
h
i
j
k
l
10µm

Plate 3 Mitotic metaphase and meiosis in diploid and triploid *Crataegus* taxa, all to the same scale. *C. punctata*, mitotic metaphase: (a) OTU 514 and (b) OTU 519. *C. ?grandis*, mitotic metaphase: (c) OTU 601 and (d) OTU 602. 20-stamen *C. crus-galli*, mitotic metaphase: (e) OTU 105, p-dichlorobenzene pretreatment, and (f) OTU 412. Meiosis in Colchicine pretreated material of probable triploid 10-stamen *C. crus-galli*, OTU 755: (g) pachytene; (h, i) diakinesis, showing nucleoli (N) and chromatic strands (arrows) between chromosomes or groups of chromosomes; and (j) metaphase. See text and Table 9.1 for additional details.


a
c
e
b
d
f
h
g
i
j
10µm

Table 9.1 Chromosome Data for Ontario individuals of Crataegus crus-galli L. sensu lato and Crataegus punctata Jacq. examined in the present study.

OTU	Site	Taxon	Collection (1)	(2)	(3)	(4)	Count (5)
105	1	C. crus-galli s. l., 20-stamen	CR387-L	20. 4. 80	*	none	3X
			CR399-L	20. 4. 80	*	PDCB	3X
			CR541-F	10. 5. 80	115	Clch.	MP
111	1	" 10-stamen	CR499-L	1. 5. 80	57	none	4X
			CR504-L	2. 5. 80	*	Clch	4X
			CR506-L	2. 5. 80	*	PDCB	4X
123	1	" "	CR500-L	1. 5. 80	57	none	4X
			CR510-L	2. 5. 80	*	PDCB	4X
131	1	" 20-stamen	CR412-L	20. 4. 80	*	Cold	3X

=Phipps 4719, 3X (Muniyamma and Phipps, 1979b)

Table 9.1 Cont.

OTU	Site	Taxon		Collection (1)	(2)	(3)	(4)	Count (5)
132	1	C. crus-galli s. l.,	20-stamen	CR421-L	20. 4. 80	*	Cold	3X
134#	1	"	"	CR416-L	20. 4. 80	*	Cold	3X
				CR533-F	8. 5. 80	115	none	>5X TP
302	3	"	10-stamen	CR525-L	6. 5. 80	107	Clch	4X
				CR872-L	27. 4. 81	*	Cold	4X
307	3	"	"	CR877-L	27. 4. 81	*	Cold	4X
401	2	"	"	CR828-L	16. 4. 81	*	Cold	4X
412	2	"	20-stamen	CR427-L	24. 4. 80	38	none	3X
417	2	"	"	CR430-L	24. 4. 80	38	none	3X

Table 9.1 Cont.

OTU	Site	Taxon	Collection (1)	(2)	(3)	(4)	Count (5)
418	2	C. crus-galli s. l., 20-stamen	CR830-L	16. 4. 81	*	Cold	3X
424	2	" 10-stamen	CR831-L	16. 4. 81	*	Cold	4X
514	5	C. punctata	PU74-L	10. 4. 80	*	Cold	2X
519	5	"	PU66-L	8. 4. 80	*	Cold	2X
601	4	C. ?grandis	CR382-L	13. 4. 80	*	none	3X
602	4	"	CR386-L	13. 4. 80	*	Cold	3X
606	4	C. crus-galli s. l., 10-stamen	CR913-F	13. 5. 81	138	none	4X
610	4	" "	CR821-L	15. 4. 81	*	Cold	4X

Table 9.1 Cont.

OTU	Site	Taxon	Collection (1)	(2)	(3)	(4)	Count (5)
622	4	C. crus-galli s. l., 10-stamen	CR824-L	15. 4. 81	*	Cold	4X
714	9	" 20-stamen	CR854-L	19. 4. 81	72	none	4X
721	12	" 10-stamen					
		=Phipps 4637, 4X (Muniyamma and Phipps, 1979b)					
722	12	" 10-stamen					
		=Phipps 4607a, 3X (Muniyamma and Phipps, 1979b)					
731	10	" 10-stamen	CR885-L	2. 5. 81	94	none	4X
		=Phipps 4598, 2X (Muniyamma and Phipps, 1979b)					
751	11	" 20-stamen	CR834-L	16. 4. 81	*	Cold	4X

Table 9.1 Cont.

OTU	Site	Taxon	Collection (1)	Collection (2)	Collection (3)	Collection (4)	Count (5)
752	2	C. crus-galli s. l., 20-stamen	CR218-F	13. 5. 79	150	Clch	4X
			CR246-F	15. 5. 79	163	Clch	MP,MM,T
			CR832-L	16. 4. 81	*	Cold	4X
753	2	" "	CR833-L	16. 4. 81	*	Cold	4X
754	2	" "	CR889-F	2. 5. 81	94	none	3X
755#	2	" 10-stamen	CR5-F	11. 5. 78	*	Clch	MP,MM,T
			CR18-F	25. 5. 78	*	Clch	MP,MM,T
			CR18-F	"	*	"	?3X
801	6	C. sp. aff. C. bushii	CR484-F	30. 4. 80	*	none	4X

Table 9.1 Cont.

(1) Collection number; L = leaf buds, F = flower buds.

(2) Fixation date.

(3) Accumulated heat (London °days C) at time of fixation; * = material forced indoors prior to fixation.

(4) Pretreatment: Clch = Colchicine 0.3 - 0.33%; Cold = incubation at 5° C; PDCB = p-dichlorobenzene, saturated solution.

(5) X = 17. Counts are of somatic cells only. MP = meiotic prophase; MM = meiotic metaphase; T = tetrads of microspores; TP = tapetal cell.

= OTU not included in morphometric studies.

Table 9.2 Numbers of diploid, triploid and tetraploid chromosome counts for OTUs in taxa comprising Ontario Crataegus crus-galli L. sensu lato. These include six euploid counts reported by Muniyamma and Phipps (1979b) in addition to the 25 reported here for the first time (Table 9.1). Numbers of sites represented are indicated in parentheses.

	2X	3X	4X	Total
10-stamen C. crus-galli	1a	3(3)	11b(6)	14c
20-stamen C. crus-galli	0	8(2)	5(4)	13
C. sp. aff. C. bushii	0	0	1	1
C. ?grandis	0	2(1)	0	2
Total	1a	13	17b	30

(a) Diploid count for OTU 731 reported by Muniyamma and Phipps (1979b).

(b) Includes tetraploid count for OTU 731 obtained in the present study (Table 9.1).

(c) OTU 731 counted only once.

9.4 Discussion

The difficulties reported by Muniyamma and Phipps (1979b) in obtaining chromosome counts from somatic material were avoided here by the use of shoot apices and leaf primordia rather than roots. Further work would undoubtedly result in a more effective feulgen staining schedule. Alternatively, in view of Longley's comment on the affinity of hematoxylin for the chromosomes of Crataegus section Crus-galli (Longley, 1924), use could be made of the hematoxylin squash staining schedules that have been published (e.g. Nuñez, 1968).

The only previous studies of the chromosome numbers in Crataegus section Crus-galli are those of Longley (1924) and Muniyamma and Phipps (1979b). These are also the only studies to have reported chromosome numbers for C. punctata. Longley used sections of PMCs to make his counts, and arrived at a base number of x = 16 for the genus. This was subsequently corrected by Moffett (1931), who determined the base number for the entire subfamily Maloideae to be x = 17, using squash techniques and material of one or more species of each of fourteen different genera and subgenera. The base number of 17 for the subfamily has been maintained by all subsequent investigators (Gladkova, 1968; Calder, Tayor & Mulligan, 1968; Byatt & Murray, 1977; Muniyamma & Phipps, 1979b), as well as by the results obtained here.

Longley's determinations of the ploidy levels in the material he examined are unlikely to be seriously affected by his error in determining the base number. The material he used was collected from individuals growing at the Arnold Arboretum that had been determined by C. S. Sargent as representing 82 "species". These included six entities in section Crus-galli and four in section Punctatae. Of the former, only an individual of C. canbyi Sarg. was found to be a diploid, as well as pollen fertile. The remaining entities represented C. crus-galli L., C. palmeri Sarg., C. reverchoni Sarg. var. discolor (Sarg.) Palmer and C. fontanesiana (Spach) Steud. and were described as partially male-sterile triploids (Longley, 1924). Crataegus punctata was determined to be a diploid; both diploid and triploid individuals were found in the related C. collina Chapm., while one individual of C. punctata var. pausiaca (Ashe) Palmer was found to be triploid.

Muniyamma and Phipps (1979b) examined seven individuals of C. crus-galli s. l. and reported one diploid, three triploids, two tetraploids and one near tetraploid aneuploid (2n = 64), based largely or entirely on PMC meioses. The individual scored as a diploid (Phipps 4598 = OTU 731) was relocated in 1977 and determined in the present study to be a tetraploid, based on somatic material (Plate 2i; Table 9.1). In view of the predominant use of meiotic material in

the earlier study, and the fact that "Even putative diploids were thus not necessarily easy to count" (Muniyamma & Phipps 1979b), it seems likely that in this case the diploid determination is incorrect, probably due to misinterpretation of multivalents as bivalents.

Neither higher stamen number nor large flowers or fruit is consistently associated with tetraploids as opposed to triploids in C. crus-galli s. l.. Both ploidy levels occur within 10- and 20-stamen C. crus-galli without appreciable morphological differentiation (Fig. 11.1; see also Section 10.2.2 for the results of comparisons of pollen grain size). In 20-stamen C. crus-galli triploids and tetraploids are distinguished from each other by the pollen infertility of the former (Section 10.2.3). Tetraploid 20-stamen C. crus-galli OTUs also were distinguished on the basis of flower and fruit descriptors in some of the cluster analyses (Chapter 5) and ordinations (Fig. 11.1; Section 11.4.2).

The apparent absence of diploid C. crus-galli s. l. in Ontario resembles the situation in Ontario C. pruinosa s. l. and C. section Tenuifoliae (Muniyamma & Phipps, 1979b) where only triploids and tetraploids were reported. In the only other section to be investigated intensively to date, C. section Rotundifoliae, the only diploids discovered to date are four reported by Muniyamma and Phipps (1979b). In a large number of counts also made exclusively from somatic

material P. G. Smith (personal communication) has found only triploids and tetraploids, predominantly the latter. There appear to be no reports as yet of chromosome numbers for any of these series from elsewhere in their range, apart from those of Longley (1924). As a result it is impossible to say whether or not Crataegus taxa found to be polyploid in Ontario are represented by diploids in the southern parts of their range which were never glaciated. Such patterns of distribution of different ploidy levels within species and genera have been recorded in a number of North American taxa (Lewis, 1980; Bayer & Stebbins, 1981), and might be expected in Crataegus species as well.

Further discussion of polyploidy in C. crus-galli s. l. will be deferred to Chapter 11, where the data reported here can be integrated with those from other chapters. An integral part of such a discussion is consideration of the means by which polyploidy may originate; some of these are described for C. crus-galli s. l. in the following chapter.

CHAPTER TEN

REPRODUCTIVE PHENOMENA

10.1 Introduction

The analyses in Chapters 5-8 have documented the morphological distinctness of the seven topodeme samples. They demonstrated also the high degree of internal homogeneity of the six samples of Crataegus crus-galli s. l. (T1-T6), in contrast to the relative heterogeneity of the sample of C. punctata (T7; Section 7.5; Section 8.7; Fig. 8.7). These analyses also showed the intermediacy of 20-stamen C. crus-galli with respect to the 10-stamen morphotype and C. punctata (Section 8.9.5; Fig. 8.4), and of the OTUs of C. sp. aff. C. bushii with respect to those of C. section Macracanthae and 10-stamen C. crus-galli (Fig. 6.2a). Crataegus punctata has been found to be exclusively diploid in Ontario, whereas C. crus-galli s. l. appears to be a polyploid complex (Chapter 9; Muniyamma & Phipps, 1979b).

One possible explanation for both the uniformity of topodemes of C. crus-galli s. l., and of the occurrence of polyploidy, is the phenomenon of apomixis. Apomixis in general is the production of offspring without gamete fusion

taking place. It may be subdivided into agamospermy and vegetative reproduction according to whether or not the process leads to seed formation (Fagerlind, 1940; Nygren, 1967). In the former case ovule development is modified, in some cases (apospory, diplospory) so that an unreduced gametophyte is produced. In these cases seed development usually follows parthenogenetic development of the unreduced gametes or other cells of the gametophyte into an embryo sporophyte. Processes such as agamospermy represent a means by which large numbers of offspring may be obtained that are identical to the parent individual, as recognized by Palmer (1932). Moreover, fertilization of unreduced gametes will lead to offspring of a higher ploidy level (Camp & Gilly, 1943; Harlan & deWet, 1975). Although vegetative reproduction (root-budding) does occur to a limited extent in some Crataegus species, its role in dispersal and the establishment of new topodemes is so limited that it is not considered further here.

The occurrence of hybridization is a possible explanation of the intermediacy of groups of OTUs such as those of C. sp. aff. C. bushii or the 20-stamen C. crus-galli ones. Although the studies described below do not bear directly on the question of hybridization, the significance of this process will be discussed in Section 10.6 and in Chapter 11.

Sections 10.2, 10.3 and 10.5 deal with three reproductive phenomena of interest for an understanding of the circumstances under which seed are produced in C. crus-galli s. l. and C. punctata. The first of these reports the results of examining the amount and quality of pollen produced, in relation to the different levels of ploidy present in these taxa (Chapter 9). The second documents the potential for agamospermy in C. crus-galli s. l. and C. punctata. Section 10.5 presents the results of field experiments designed to determine what pollinations, if any at all, are necessary for seed-set to occur. The results of observations on floral visitors, especially pollinators, are reported in Section 10.4. References will be made throughout the chapter also to additional data on reproductive morphology in Appendix 3.

The concluding section of this chapter discusses these and related results from the standpoint of their significance for the reproductive success of Crataegus individuals. Having done so, their significance for the origin of topodemes and for the evolution of taxa can be discussed separately, in Chapter 11.

10.2 Androecial Development, Pollen Size and Pollen Production

Stamens develop rapidly from apparently undifferentiated primordia when growth resumes after the winter (Appendix 3, Plate 7a-e). In both C. crus-galli s. l. and C. punctata pollen meiosis occurs well before flowers open (Section 9.3); at a stage when the ovules consist of little more than the primordia of the nucellus and integuments (Plate 4a, b; 5a). Pollen of both taxa appears to be mature prior to anthesis (Plate 4d). In both taxa anther structure resembles that described by Davis (1966) for the Rosaceae as a whole, with respect to the following: persistent epidermis; endothecium with fibrous thickenings; and multinucleate glandular tapetum (Plate 4a-d).

Pollen size was studied in order to see whether or not it was independent of ploidy level. Pollen morphology was not examined in view of the high degree of variability in exine pattern found within European Crataegus taxa by Byatt (1976c).

Quantitative variation in pollen production was investigated because of the relevance of this parameter to characterization of the breeding system (Cruden, 1977; Cruden & Miller-Ward, 1981), and the variation in stamen

number found in C. crus-galli s. l. Variation in pollen quality was studied because of its significance for the interpretation of the pollination experiments described in Section 10.5 below.

These pollen studies were carried out on a deterministically selected subsample of the total 160 OTU sample, plus a limited number of additional individuals. This sample included all of the OTUs involved in the experiments discussed in Section 10.5 plus additional OTUs, from each topodeme random sample, and from selected additional sites (Appendix 4).

10.2.1 Assessing Pollen Size, Production and Fertility: Method

At anthesis unopened and open flowers were fixed in Farmer's fluid (Section 9.2) or other fixatives described below (Section 10.3.1). In scoring stamen number in unopened flowers, undehisced anthers were set aside, stored in 70% ethanol, and subsequently used to estimate pollen stainability and number of pollen grains per anther.

Pollen size was measured directly using material prepared in the course of estimating pollen quantity and quality. Pollen grains containing cytoplasm (see below) were mounted in glycerine jelly (Sass, 1958) and measured

using an eyepiece micrometer scale. In order to avoid variation in the extent to which grains swelled, all measurements were made 20 minutes after mounting in glycerine jelly (Faegri & Iversen, 1964; Stanley & Linskens, 1974). Two perpendicular measurements of pollen grain diameter were made and averaged. These averages were calculated for 30-50 pollen grains for each OTU, and used to estimate OTU mean pollen diameter. They were also used to estimate the frequency of 0.1 eyepiece unit size classes among the three ploidy levels in the sample. The association of pollen grain size and ploidy level was assessed by means of G-statistic analyses of these frequencies (Sokal & Rohlf, 1969).

Pollen stainability was determined for an individual tree by macerating fifty anthers overnight in 0.25 ml of a Malachite Green - Acid Fuchsin - Orange G mixture (Alexander, 1969). The next day 0.25 ml glycerine jelly was added to the stain. The resulting volume permitted 15 hematocytometer determinations of pollen grain concentration, based on direct counts of a total of 2000-3000 pollen grains per OTU. The results of the 15 counts were pooled to estimate the total number of pollen grains per anther. In addition, the pollen grains were classified according to whether cytoplasm was present and stained with Acid Fuchsin and Orange G or was absent, in

which case only Malachite Green was absorbed by their walls. The proportion of triple-stained pollen grains was used to estimate the total number of stainable pollen grains per anther. This number was used together with mean stamen number per flower and mean style number per flower to calculate tree pollen-ovule ratios (P/Os), on the basis of two ovules per style.

10.2.2 Results - Pollen Size

Pollen grains of C. crus-galli s. l. are tricolporate and were found to have polar diameters in the range 38.4-40.6 µm, and equatorial diameters in the range 36.9-48.7 µm (Table A4.1, Appendix 4). In view of their oblate spheroidal shape, polar views of pollen grains were much more frequent than equatorial ones. In the case of C. punctata only equatorial diameters were measured, for two OTUs. For these the mean diameter was found to be 36.0 and 38.1 µm (Table A4.1, Appendix 4). The two individuals of C. sp. aff. C. bushii studied in Section 10.5 had mean diameters of 44.9 and 45.0 µm (Table A4.1, Appendix 4).

A null hypothesis of the independence of pollen grain diameter and ploidy level was rejected (Table 10.1a). Further testing suggested that this result is primarily due to the difference between C. crus-galli s. l. (3x, 4x) and C. punctata (2x; Table 10.1b, c). The significant value of

G obtained comparing pollen diameter class frequencies among 3x and 4x OTUs of C. crus-galli s. l. (Table 10.1c) appeared to reflect the difference between the two frequency distributions rather than an appreciable difference in diameter between these two ploidy levels.

Table 10.1 Tests of the independence of pollen grain diameter and ploidy level. Test criterion is the G-statistic, which is compared with χ^2 for the degrees of freedom indicated (Sokal & Rohlf, 1969). Entries in tables (a-c) are numbers of pollen grains in each diameter class.

Pollen grain diameter (equatorial plane) classes in eyepiece micrometer units (1 unit = 32 µm).

	1.0	1.1	1.2	1.3	1.4	1.5	Total
(a) 4x (N = 7 OTUs)	0	29	46	63	76	43	257
3x (N = 4 OTUs)	0	0	22	56	43	26	147
2x (N = 2 OTUs)	16	38	25	4	0	0	83
Total	16	67	93	123	119	69	487

d.f. = 10 G = 241.93***

(b) crus-galli (3x, 4x)	0	29	68	119	119	69	404
punctata (2x)	16	38	25	4	0	0	83
Total	16	67	93	123	119	69	487

d.f. = 5 G = 209.49***

Table 10.1 Continued

Pollen grain diameter (equatorial plane) classes in eyepiece micrometer units (1 unit = 32 μm).

	1.0	1.1	1.2	1.3	1.4	1.5	Total
(c) 4x *crus-galli*	0	29	46	63	76	43	257
3x *crus-galli*	0	0	22	56	43	26	147
Total	0	29	68	119	119	69	404

d.f. = 5 G = 32.44***

*** $p < 0.001$

10.2.3 Results - Pollen Stainability

Pollen stainability is high in all but one or two of the individuals of C. punctata studied (Fig. 10.1; Table A4.2, Appendix 4). Within C. crus-galli s. l., however, pollen stainability varied appreciably (Table A4.2, Appendix 4). The tetraploid individuals of the 10-stamen morphotype (Table 9.1) had uniformly high pollen stainability, in excess of 60% (Fig. 10.1). This was also true of C. sp. aff. C. bushii (Site 6), as well as tetraploid individuals of 20-stamen crus-galli (Table 9.1). Other individuals of the 20-stamen morphotype which have been determined to be triploids, and one of the 10-stamen triploids (OTU 755; Table 9.1), have pollen stainabilities in the range 11-37% (Fig. 10.1). All of the individuals of C. ?grandis examined at Site 4 are uniformly male-sterile, with very small anthers containing no pollen at all.

The value of pollen stainability as an indicator of the presence of cytoplasm in pollen grains, and hence of the potential for pollen function is borne out for the Crataegus material examined here by the pollination experiments (Section 10.5). It is also indicated by the association found by Byatt, Ferguson and Murray (1977) between positive results in tests for the presence of an intact vegetative cell plasma membrane (using fluorescein diacetate, Heslop-Harrison & Heslop-Harrison, 1970) and similar results

in tests based on staining pollen grain cytoplasm with cotton blue in lactophenol.

10.2.4 Results - Pollen-ovule Ratios

Individuals of C. crus-galli s.l. produced a total of 10,000 to 30,000 pollen grains per anther regardless of morphotype. Individuals of C. punctata showed a slightly higher range (15,000 - 35,000). Being based on stainable pollen grains, however, P/O was strongly associated with pollen stainability (Fig. 10.1; Table A4.2, Appendix 4), with patterns of relative magnitude similar to those described above; highly pollen-fertile individuals tended to have high P/Os. Individuals with low P/Os tended to be the pollen-infertile 20-stamen crus-galli (Fig. 10.1).

Figure 10.1 Pollen-ovule ratios and pollen stainability in *Crataegus* taxa:

20-stamen *C. crus-galli* (open symbols);

10-stamen *C. crus-galli* (solid symbols);

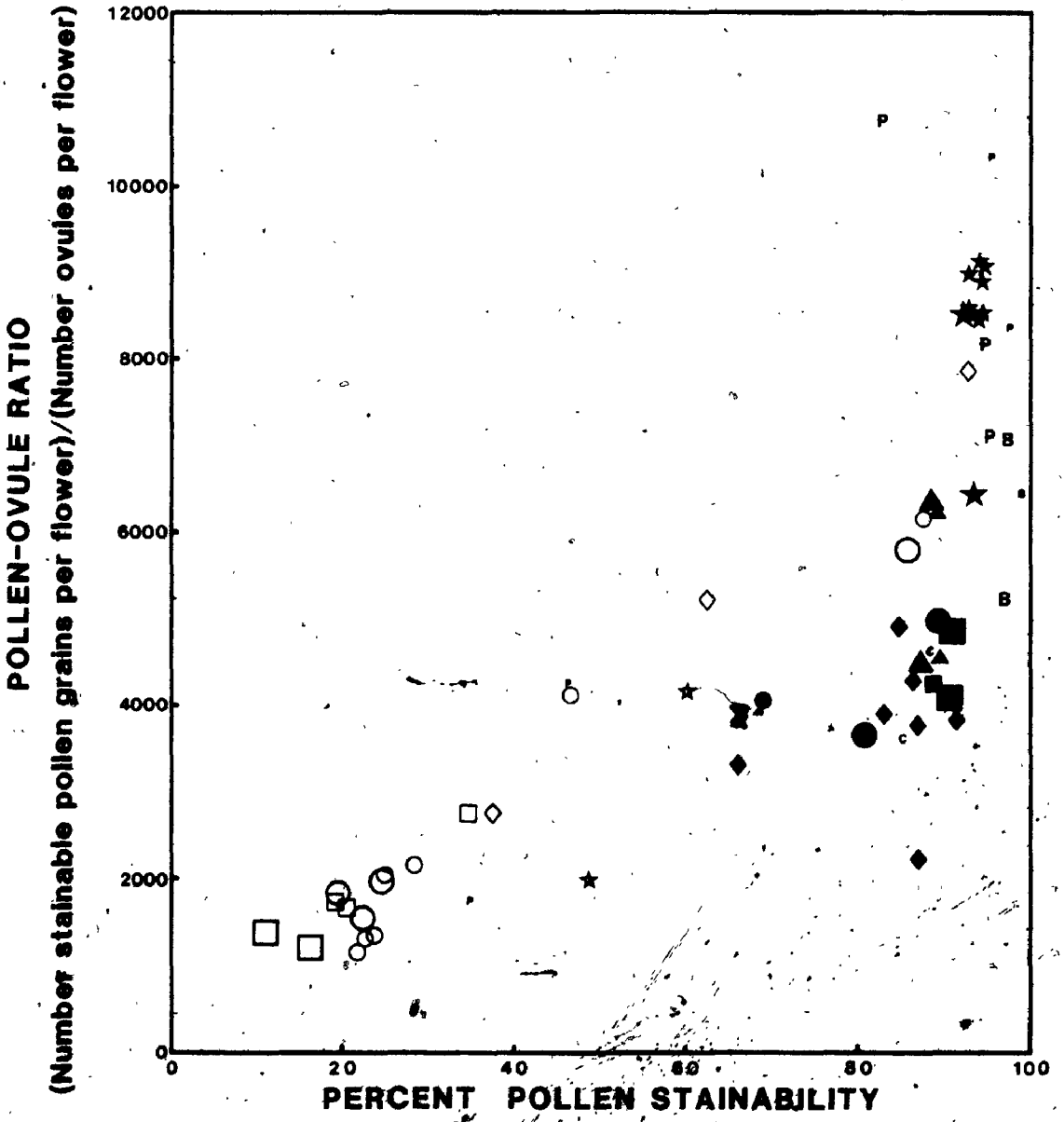
C. sp. aff. *C. bushii* (Site 6, B);

C. punctata (Site 5, P);

C. conspecta Sarg. (Site 4, C);

and individuals intermediate between 10-stamen *C. crus-galli* and *C.* ?*grandis* (Site 4, half-solid stars).

Sites are abbreviated here as follows: Site 1, squares; Site 2, circles; Site 3, triangles; Site 4, stars; and additional sites not otherwise specified, diamonds. The larger symbols indicate individuals studied experimentally (Section 10.5). See Appendix 4 and text for details.


POLLEN-OVULE RATIO
(Number stainable pollen grains per flower)/(Number ovules per flower)
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100
PERCENT POLLEN STAINABILITY
P
P
P B
B

10.3 Gynoecial Development

Within the gynoecium only ovule development was studied, in order to determine whether or not seed production in C. crus-galli s. l. and C. punctata was associated with non-sexual processes. Both apospory and diplospory have been reported in other Crataegus species (Muniyamma & Phipps, 1979a and unpubl.).

The following section describes general methods used for liquid preservation of floral material including not only that destined for use in studying ovule development but also that used in documenting the morphology of mature flowers (Section 3.2, 3.3; Appendix 3, Plate 8) and their early development (Appendix 3, Plate 7).

10.3.1 Methods

Floral material ranging from bisected, unexpanded inflorescence buds to mature flowers at or shortly past the time of anthesis was collected into vials of fixative chilled on ice, usually in the field, but sometimes also in the lab, from cut branches forced indoors. In every case the collection date was also characterized by the corresponding temperature sum (° days of vernal heat accumulation; Chapter 9; Appendix 7).

A range of fixatives was employed, from general purpose ones such as Farmer's fluid (Section 9.2) and formalin, propionic acid, 50% EtOH 5:5:90 (FPA_{50}) to special purpose ones such as Nawaschin (Craf V; Berlyn & Miksche, 1976) and acrolein-based fixatives (Feder & O'Brien, 1968; Cheng, Greyson and Walden, 1979). Material fixed in FPA_{50} and the Craf V was immediately aspirated with a hand vacuum pump to remove air from the tissues and ensure rapid penetration of the fixatives. For the same reason the hypanthia of mature flowers were usually slashed with a razor blade before being put in any of the fixatives used. Similarly, small amounts of detergent (TWEEN-20) were sometimes incorporated into the Craf V and acrolein fixatives to improve penetration.

Material was fixed for 24-48 hours. The fixative was then discarded and replaced with 70% EtOH (Farmer's and FPA_{50} fixed material) or distilled water (plus a chrystal of thymol to inhibit the growth of mold; Craf V and acrolein fixed material). The material was then stored in the cold until used.

Observations of floral development and gross floral morphology were made using mainly Farmer's and FPA_{50} fixed material. Scoring floral descriptors for the numerical studies (Chapter 3) was done using unstained material in a dish of 70% EtOH, under a dissecting microscope. Stages of floral development (Appendix 3, Plate 7) were photographed

with Leitz Ultropak dipping cone objectives, using material stained in acid fuchsin (0.5% in 95% EtOH), dissected out and examined in absolute EtOH (Sattler, 1968; Posluszny, Scott and Sattler, 1980). Mature stigmatic surfaces and ovules (Appendix 3, Plate 8) were photographed with a Nikon SMZ-10 dissecting microscope, also using acid fuchsin stained material.

Observations on ovule and embryo-sac development were made on sectioned material. All material except that fixed in Craf V was dehydrated by standard methods employing a graded EtOH-Tertiary Butyl Alcohol (TBA) series (Berlyn & Miksche, 1976). The latter material was glycerine dehydrated (Berlyn & Miksche, 1976); the glycerine was then replaced progressively with absolute EtOH (Cheng, Walden & Greyson, 1979). From absolute EtOH the material was transferred to the EtOH-TBA series.

From TBA material was transferred to a 1:1 mixture of TBA and the embedding medium (Tissueprep, m.p. 61.0 C; Fisher) and then to two changes of pure medium before being embedded. Sections 7-10 μm thick were cut from blocks of embedded material using an A-O Spencer rotary microtome and either a standard knife (A-O Spencer) or disposable stainless steel blades in a special holder (Feather Industries Ltd.).

Ribbons of sections were affixed to slides either with Haupt's adhesive or, to enable Feulgen staining, using a gelatin-chrome alum adhesive (Berlyn & Miksche, 1976). Sections were stained using standard methods and the following stain combinations (Clark, 1973): safranin O - fast green FCF; safranin O - aniline blue W.S.; and Johansen's hematoxylin - erythrosin. The hematoxylin schedule used involved mordanting sections in 4% aqueous $NH_4Fe(SO_4)_2 \cdot 12H_2O$ containing 1% glacial acetic acid and 0.1% concentrated sulphuric acid. Sections were overstained and then destained in the mordant solution diluted 1:1 with distilled water (Clark, 1973). Following dehydration and clearing, sections were mounted in Permount (Fisher).

The following description of ovule development is based on observations made on sectioned material of both C. crus-galli s. l. and C. punctata, since events in both taxa are largely similar. The OTUs selected for study were mainly those on which the pollination experiments described in Section 10.5 were performed. While mainly lower ovules are represented in the illustrations in Plates 4 - 6, virtually all of the features observed in lower ovules were observed in upper ones also.

10.3.2 Results - Ovule and Embryo-sac Development

Ovule primordia arise on the inner (adaxial) surfaces of each gynoecial primordium sometime after the free portions of the latter begin to develop into the styles (Appendix 3, Plate 7e). By the time the stigmatic surface has assumed the papillate condition exhibited at maturity (Appendix 3, Plate 8b; Plate 4e) and pollen meiosis occurs (Plate 4a, b) the two superposed ovules in each locule consist of a nucellar primordium surrounded by the primordia of the inner and outer integuments (Plate 5a).

Ovules are crassinucellate as in other Rosaceae with bitegmic ovules (Plate 4, 5; Davis, 1966). A multicellular archesporium develops as a result of periclinal divisions in the subepidermal layer of the nucellus (Plate 5a, b; Plate 4f; Liljefors, 1953). Megaspore mother cell differentiation was observed infrequently, but presumably occurs in a single archesporial cell, as in Sorbus and other Rosaceae (Liljefors, 1953; Davis, 1966). An obturator develops at the attachment of the two funiculi to the gynoecial primordium (Plate 4f; Appendix 3, Plate 8a, c, d; Péchoutre, 1902).

Eight-nucleate Polygonum-type embryo-sacs (Plate 5c) arise either from one of the megaspores or from one or more unreduced archesporial cells, following degeneration of the

megaspore mother cell or of the megaspores (Liljefors, 1953; Muniyamma & Phipps, 1979a). The antipodals degenerate and at anthesis mature embryo-sacs usually consist of only the egg apparatus (synergids and egg) and the central cell (Plate 4g; Plate 5d).

In open flowers numerous granules are present in some of the mature embryo sacs (Plate 4g). These may correspond to the starch granules found in the embryo-sacs of Spiraea (Davis, 1966) or they may be indicative of endosperm development. Endosperm in C. crus-galli s. l. (C. ?grandis, Plate 6g-i; 10-stamen C. crus-galli, not shown) and possibly also C. punctata (Plate 4g) appears to involve formation of a haustorium that reaches the chalaza (Péchoutre, 1902), as described by Davis (1966) also for Prunus and Rubus. Embryogeny was not studied, although embryos were observed occasionally (Plate 6h).

10.3.3 Results - Evidence of Agamospermy

In this study the principal evidence for the occurrence of agamospermy was the observation of more than one embryo-sac within a single ovule. Sexual reproduction in flowering plants almost invariably involves development of a single female gametophyte from one or more of the megaspores produced by megaspore mother cell meiosis. Agamospermy through diplospory would differ only in that the resulting

gametophyte would contain nuclei with the unreduced number of chromosomes, and would be able to develop an embryo parthenogenetically. Diplospory would be detectable anatomically only by observation of the nuclear divisions leading to embryo-sac formation (Gustafsson, 1946). Agamospermy by means of adventitious embryogeny is frequently associated with polyembryonic seed, and in this case is usually characterized by the development of multiple embryos outside any embryo-sacs that may be present (Gustafsson, 1946). Only as a result of unreduced cells other than the megaspore mother cell developing into female gametophytes (i.e. apospory) are multiple embryo sacs likely to be produced. One or more unreduced embryo-sacs may develop in this way either in addition to or in place of a sexual embryo-sac (Davis, 1966; Muniyamma & Phipps, 1979a).

Accordingly, while diplospory in the material studied above may have escaped detection, the occurrence of multiple embryo-sacs (Plate 5, 6) is evidence for the occurrence of agamospermy through apospory.

Development of multiple embryo-sacs in the material studied here frequently appeared to be associated with the degeneration of the megaspores or megaspore mother cell, so that the aposporous embryo-sacs which developed probably did so from the surrounding unreduced archesporial tissue.

Therefore, the meristematic cells associated with the degenerated mass of the megaspores or megaspore mother cell, or in general, with the absence of a single, well-developed embryo-sac by the time anthesis occurred have been interpreted as aposporous initials (Plate 5e; Plate 6a).

The occurrence of different conditions of ovule development associated with sexual development of the female gametophyte on the one hand, and with aposporous embryo-sac development on the other, is summarized in Table 10.2 for some of the different groups of OTUs in the present study, notably those for which experimental data is also available (Section 10.5).

It is evident from the data reported in Table 10.2 that apospory, as indicated by the presence of multiple embryo-sacs, occurs in all four morphotypes of C. crus-galli s. l. studied here, as well as in both triploid and tetraploid OTUs. It was also evident in the material examined that this phenomenon was absent in C. punctata, although with OTU 519 some ovules were found in which possible aposporous initials were present (Plate 4h; Table 10.2). Multiple embryo-sacs were also found in material of C. macracantha (Table 10.2).

In both 10- and 20-stamen C. crus-galli single embryo-sacs were also observed (Table 10.2). In the

tetraploid OTUs these could have been the result either of sexual processes only, or of either sexual or aposporic development, coupled with obliteration of any additional embryo-sacs of whatever origin that might have been present earlier. In the case of the triploid 20-stamen C. crus-galli at Site 2 the role of sexual processes in producing the single embryo-sacs observed seems less likely.

Plate 4 Androecial and gynoecial anatomy of Crataegus crus-galli s. l. and C. punctata: (a, b) pollen meiosis in 20-stamen C. crus-galli (OTU 752, 149 London ° days, 1979; (a) at the same scale as (h), (b) at the same scale as (e)); (c) C. ?grandis anther shortly before anthesis (at same scale as (e); OTU 602, 201 London ° days, 1980); C. punctata anther (d), stigma (e) and ovule (f) development shortly before anthesis (OTU 514, 186 London ° days); (g) mature embryo-sac, lower ovule of open flower, C. punctata (at same scale as (d); OTU 514, 294 London ° days, 1979); and (h) immature embryo-sac and possible aposporous initials in lower ovule of open flower, C. punctata (OTU 514, 294 London ° days. In (a, b, d-h), CRAF fixation and glycerine dehydration; in (c), Farmer's, and TBA dehydration only. All stained with hematoxylin and erythrosin. Explanation of labelling: ?Ai, possible aposporous initials; Cn, central cell nucleus; Ep, epidermis; Et, endothecium; iI, inner integument; oI, outer integument; N, nucellus; P, papillae; Pmc, pollen mother cell; Po, pollen grain; ST, stigmatoid tissue; and T, glandular tapetum. See Table 10.2 for further details including site and ploidy of the OTUs represented here.

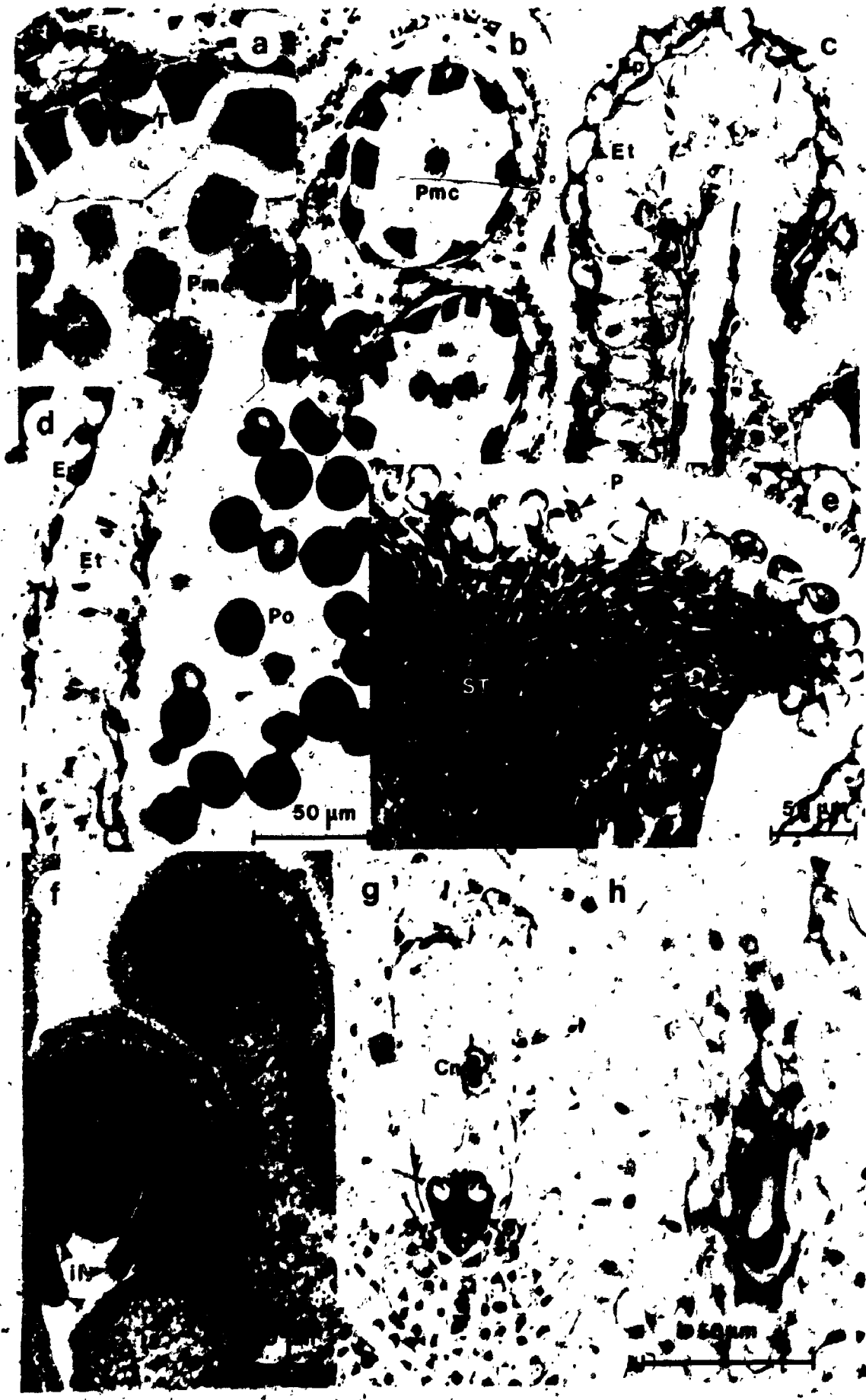

a
b
c
Et
T
Pmc
Pm
Et
d
e
E
Et
P
Po
ST
50 µm
50 µm
f
g
h
IN
Cr
0.5µm

Plate 5 Ovule anatomy in Crataegus crus-galli s. l.: early ovule development (a; upper ovule) and archesporial differentiation (b) in 20-stamen C. crus-galli (OTU 752, 150 London ° days, 1979; both to same scale); single embryo-sac in closed (c) and open (d) flower (OTU 111, 370 London ° days, 1979; (c) to same scale as (h)); and 10-stamen C. crus-galli, (e) aposporous initials and degenerated mass of megaspores or megaspore mother cell (arrow) and (f) multiple embryo-sacs in open flowers (OTU 606, 320 London ° days, 1981; both to same scale as (h)), multiple embryo sacs in (g) OTU 416 (328 London ° days, 1979) and (h) OTU 307 (upper ovule; 367 London ° days, 1980). Except as noted, only lower ovules are shown. In (a, b, d, g), CRAF fixation and glycerine dehydration; (c), acrolein fixation, glycerine dehydration; and (e, f, h) Farmer's, and TBA dehydration only. All stained with hematoxylin and erythrosin. Explanation of labelling: An, antipodals; Cn, central cell nucleus; iI, inner integument; oI, outer integument; N, nucellus; and Sy, synergids. See Table 10.2 for further details including site and ploidy of OTUs represented here.


ol
50µm
a
b
il
N
c
d
An
Cn
Sy
e
f
g
h

Plate 6 Ovule anatomy in Crataegus crus-galli s. l.: aposporous initials (a; to same scale as (d); at arrow mass of degenerated megaspores or megaspore mother cell) and multiple embryo sacs (b, c; adjacent sections, both to same scale) in closed flowers of C. sp. aff. C. bushii (OTU 801, 217 London ° days, 1979); multiple embryo-sacs in 20-stamen C. crus-galli (d, OTU 417, open flower, 328 London ° days, 1979; e, OTU 418, open flower, 341 London ° days, 1980); and multiple embryo sacs in C. ?grandis in open flower (f, at same scale as (d); OTU 601, 220 London ° days, 1979) and following fertilization and petal-drop (g-i, OTU 602, 380 days, 1980; in (g) and (h), adjacent sections at the same scale as (d); the same section is shown in (g) and (i)). In (a-d, f) CRAF fixation and glycerine dehydration; in (e, g-i) Farmer's, and TBA dehydration only. All stained with hematoxylin and erythrosin. Explanation of labelling: Ch, chalaza; En, endosperm; Ues, unfertilized 2-nucleate embryo-sac; and Z, zygote of fertilized embryo-sac. See Table 10.2 for details of ploidy and sites represented by OTUs shown here.


a
b
c
50 µm
d
50 µm
e
f
g
h
En
Ch
En
Ue
i

Table 10.2 Evidence of the occurrence and non-occurrence of apospory in material of Crataegus crus-galli s. l., C. punctata, and C. macracantha. See Table 9.1 for details of ploidy determination.

	Ploidy	Site	OTUs	Ovules with: Single ES(1)	Mult. ES(1)	Aposporous Initials only(1)
10-stamen C. crus-galli s. l.	4X	1	111, 123	+ (Plate 7c, d)	+	+
" "	"	2	401, 416, 424	+?	+ (Plate 7g)	+
" "	"	3	302, 307	+	+ (Plate 7h)	+
" "	"	4	606, 622	ND	+ (Plate 7f)	+ (Plate 7e)
" "	"	10	731	ND	+	ND
20-stamen C. crus-galli s. l.	3X	1	105	ND	+	ND
" "	"	2	412, 417, 418	+	+ (Plate 8d, e)	ND
" "	4X	2	752	+	+	ND
C. sp. aff. C. bushii	4X	6	801	ND	+ (Plate 8b, c)	+ (Plate 8a)
C. ?grandis	3X	4	601, 602	ND	+ (Plate 8f, i)	ND
C. punctata	2X	5	514, 519	+ (Plate 6g)	-	+? (Plate 6h)
C. macracantha	(2)	13	761	ND	+	ND

(1) ES = Embryo-sac.

+ indicates condition observed in material examined.

- indicates condition not observed in material examined.

ND indicates that material examined inadequate to distinguish occurrence and non-occurrence.

(2) Data unavailable, but other material determined to be 4X (Maniyamma & Phipps, 1979b).

10.3.4 Critique of the Embryological Approach

It is clear that the method employed here to determine whether or not agamospermy is possible in the OTUs under study is extremely profligate of time and effort for the sample size that can be examined. Clearing techniques and phase contrast (Herr, 1971) or Nomarski differential interference contrast (Coombs, unpubl.) microscopy may enable rapid analysis of larger numbers of ovules, in some cases. Limited trials of these methods with C. crus-galli s. l., however, suggest that in this material tannins may obscure the embryological details too much for this approach to be useful.

Moreover, while cytological methods such as sectioning or clearing ovules can certainly document the events of female gametophyte development, they can provide only a limited estimate of the frequency with which particular events occur. Progeny testing can provide much better estimates of the frequency with which processes such as apospory give rise to uniparental offspring, using controlled pollinations (Marshall & Brown, 1974). However, such testing requires marker traits whose inheritance is reasonably well understood. Unfortunately, these are generally unavailable in woody perennial genera like Crataegus, although in Sorbus a simple test system apparently has been found (McAllister & Gillham, 1980).

With open pollination, chromosome counts made on batches of offspring from chromosome-counted individuals could provide estimates of the frequency with which unreduced gametes are produced and are fertilized. Gottlieb (1977) suggests methods for inferring the level of heterozygosity, and hence the probable breeding system, from studies of protein polymorphism in such batches of offspring.

There are two further shortcomings of the cytological methods employed here that are worth noting. The first is the imprecision with which ovule and embryo-sac developmental stages are related to overall floral development and to external events by either calendar date or London temperature sum (Appendix 7). In the absence of a temperature sum determined daily for each site studied and correlated with floral development, a better method of describing developmental stage might be selected measurements of flower size prior to anthesis, and a record of the time elapsed since the flower opened, after.

The second difficulty in studying embryo-sac development, and particularly the events surrounding embryogenesis in Crataegus is that from the time an inflorescence begins to expand, an increasing proportion of its flowers are likely to be shed. Thus the ovules of fewer and fewer flowers will exhibit stages necessarily characteristic of successful seed-set. As a result caution

needs to be exercised with the significance attributed to observations of embryo-sac or ovule degeneration. This difficulty can be seen as the result of a lack of pollinator specialization in the flowers of Crataegus, and the production instead of very large numbers of flowers each containing only a small number (2-10) of ovules. In contrast, orchid flowers which are produced in relatively small numbers per individual and which are highly specialized for ensuring pollinator service also contain very large numbers of ovules. Such flowers are much more rewarding objects of study with respect to the occurrence of agamospermy (Catling, 1982).

10.4 Flower Visitors - Thieves and Pollinators

The primary attractants of Crataegus flowers for insect visitors are carbohydrates and protein in the form of nectar and pollen. The odor of the flowers is a secondary attractant that advertises their availability (Faegri & van der Pijl, 1979). Individually, each white, dish-shaped flower constitutes a relatively weak visual stimulus (Appendix 3; Power, unpubl., Appendix 5; Faegri & van der Pijl, 1979). Displayed in large numbers over a single individual (Appendix 3), however, the flowers are probably quite conspicuous to insects both visually and olfactorily.

Flower visitors were noted in the course of the pollination experiments described in Section 10.8 below. They were studied in detail during May and June 1980 at Sites 2 and 3 by Power (Power, unpubl.; Table A5.1, Appendix 5) in view of the importance of learning more about Crataegus pollination.

The only flower visitors whose bodies retained pollen, and thus might transfer it between flowers, were small bees and a Scarabid beetle (Appendix 5). The majority of the flower visitors recorded (Coleoptera and Diptera; Table A5.1, Appendix 5; Knuth, 1908; Proctor & Yeo, 1973) are effectively nectar or pollen thieves (Faegri & van der Pijl, 1979), since while they remove these substances they do not contribute to pollen transfer between flowers.

10.5 Reproductive Behavior

Reproductive behavior was studied by means of a series of experiments designed to determine the conditions required for seed-set to occur. Earlier experimental work which dealt exclusively with diploid Crataegus taxa (Bradshaw, 1971, 1975; Love & Feigen, 1978) demonstrated a requirement for cross-pollination. One question to be answered here is whether or not self-sterility is maintained in polyploid taxa such as Ontario C. crus-galli s. l. (Table 9.2). Another is whether, given the occurrence of unreduced

gametophytes in the ovules of these taxa (Table 10.2), any pollination at all is required for seed-set to take place. Finally, experiments such as those reported here can also provide some indication of what factors, if any, may be limiting seed production.

10.5.1 Sampling

The sites and taxa represented in the experiments described below are indicated in Table 10.3. All but two of the individuals on which the experiments were set up (OTUs 752, 754) at Sites 1 - 5 belonged to the random samples drawn at these sites. Likewise, with only three exceptions (Table 10.3) the cross-pollinations made were reciprocal among the individuals studied. At all of the sites selection of trees to be studied experimentally was deterministic, based on the abundance of unopened flowers.

Table 10.3 OTUs used in pollination experiments described in Section 10.5, showing taxon, ploidy (Table 9.1), site and year in which the experiments were carried out.

	Ploidy	Site	OTUs	Year
10-stamen *C. crus-galli* s. l.	4X	1	111, 123*	1979
" "	"	2	401, 424	1979
" "	(4X)	3	302, 306, 307	1979
" "	4X	4	606, 622*	1981
" "	"	10	731*	1982
20-stamen *C. crus-galli* s. l.	(3X)	1	105, 133, 134	1979
" "	3X	2	412, 418	1980
" "	"	2	754	1981
" "	4X	2	752	1981
C. sp. aff. *C. bushii*	ND	6	811, 813	1980
C. ?*grandis*	3X	4	601, 602**	1980
C. punctata	(2X)	5	512, 514, 519	1979

* Additional OTUs used as pollen sources: *Phipps* 4454 (Site 1); OTU 696 (Site 4); *Dickinson* 1036 (Site 10).
** Male-sterile morphotype; individuals of *C. conspecta* used as pollen sources.
ND Data unavailable, but see that for OTU 801 (Table 9.1).

10.5.2 Methods

The experiments performed represented tests of the following: (A) the effectiveness of pollinating agents in provoking seed production by unmanipulated flowers; (B) the ability of intact flowers to produce seed when isolated from pollinating agents; (C) the ability of flowers to produce seed when emasculated as well as isolated; (D) the response of emasculated flowers to self-pollination; and the response of emasculated flowers to pollination, using pollen from another individual of the same (E) or different (F) morphotype or species, respectively. These six tests (Treatments A-F) were performed through combination of three manipulations, namely emasculation, pollination, and isolation from pollinators:

	not emasculated	emasculated
not pollinated	B	C
pollinated	A	D (self) E (cross) F (interspecific)

not isolated: A
isolated: B - F

In all cases, treatments were done at the "popcorn" stage, immediately prior to anthesis. However, none of the flowers treated had yet opened. Each treated inflorescence

represented a replicate in the experiment on a tree, associated with a unique serial number. It was usually possible to provide a minimum of five flowers per replicate, and six replicates per treatment per tree.

Inflorescences were selected that had the maximum number of mature, unopened flowers. Each inflorescence selected was numbered and manipulated as required by the experimental protocol. Emasculations were performed on unopened flowers as close to anthesis as possible, using watchmaker's forceps to remove the stamens. The number of styles in each emasculated flower was recorded. Flowers with evident insect or fungal parasites, or in which anther dehiscence had already begun were rejected and removed. All inflorescences regardless of treatment (i.e. A - F) were also culled of all flowers already open or too small for manipulation. Emasculated flowers were pollinated by dabbing the contents of newly dehisced anthers onto the stigmas. Individuals between which cross-pollinations were made were 30 to 240 m apart. Inflorescences were isolated using fine mesh (openings 0.4mm x 0.9mm) polyester bags approximately 10cm x 20cm, tied firmly in place with string. Observations on pollinators were made during the course of the experimental manipulations.

The inflorescences in Treatment A were also isolated with bags, 3-6 weeks after anthesis, in order to avoid loss of fruit and ensure comparable conditions for fruit maturation in all treatments. Fruit were harvested from treated inflorescences in September and October together with a sample of fruit from untreated inflorescences.

Pyrenes were extracted from the fruit and opened in order to score seed-set. A seed score was calculated for each replicate in an experiment as the total number of seed produced, plus half the number of pyrenes in which seed development had evidently begun but had not been completed, for whatever reason. Thus pyrenes in which insect larvae had developed within the testa were scored as 0.5, as were those in which the testa remained shriveled at maturity. These situations were scored in this way since they represented an intermediate outcome of the experiment (seed development begun but not completed) not necessarily related to the test being made. Fruit in which seed development was affected by Cedar Apple Rust were excluded from all calculations. From the seed score the proportion of gynoecial units (number of styles) producing seed in each replicate was calculated. For treatments A and B the number of styles was estimated from the mean number of styles per flower determined in Treatments C-E.

In order to obtain variates whose distribution more nearly approximated normality the proportion of gynoecial units producing seed was arcsin-square root transformed prior to calculating the mean and variance for each treatment (Sokal & Rohlf, 1969). The resulting statistics were then back-transformed to percent for tabulation and comparison. In the case of the untreated fruit, the seed score was calculated as the proportion of gynoecial units (pyrenes) producing seed, per fruit.

The effect of the experimental manipulations on fruit development as distinct from seed development was noted by keeping a record of the number of parthenocarpic fruit encountered in the treatment replicates.

Intact seed extracted from experimental and open-pollinated fruit were germinated following stratification to break their dormancy (Fordham, 1960).

10.5.3 Results

The results of the experiments described here (Appendix 6) are summarized in Figure 10.2 (10- and 20-stamen C. crus-galli) and Figure 10.3 (C. ?grandis; C. sp. aff. C. bushii; C. punctata).

Figure 10.2 Experimental seed-set in 10- and 20-stamen Crataegus crus-galli: (a) 10-stamen morphotype, N = 10 OTUs; (b) 20-stamen morphotype, N = 5 triploid OTUs (experiments of 1979, 1980); (c) comparison of triploid and tetraploid 20-stamen OTUs at Site 2 (1981). In (c) results on the left (mean underlined) are for the tetraploid; those on the right are for the triploid; cross-pollinations (Treatment E) were between these two OTUs. Vertical lines: range. Horizontal lines: mean. Hatched rectangle: 95% confidence interval for mean. Asterisks indicate means whose 95% confidence interval does not include zero. Treatments A - F as described in Section 10.5.2. See Table 10.3 for identification of OTUs; see Appendices 4 and 6 for additional details.

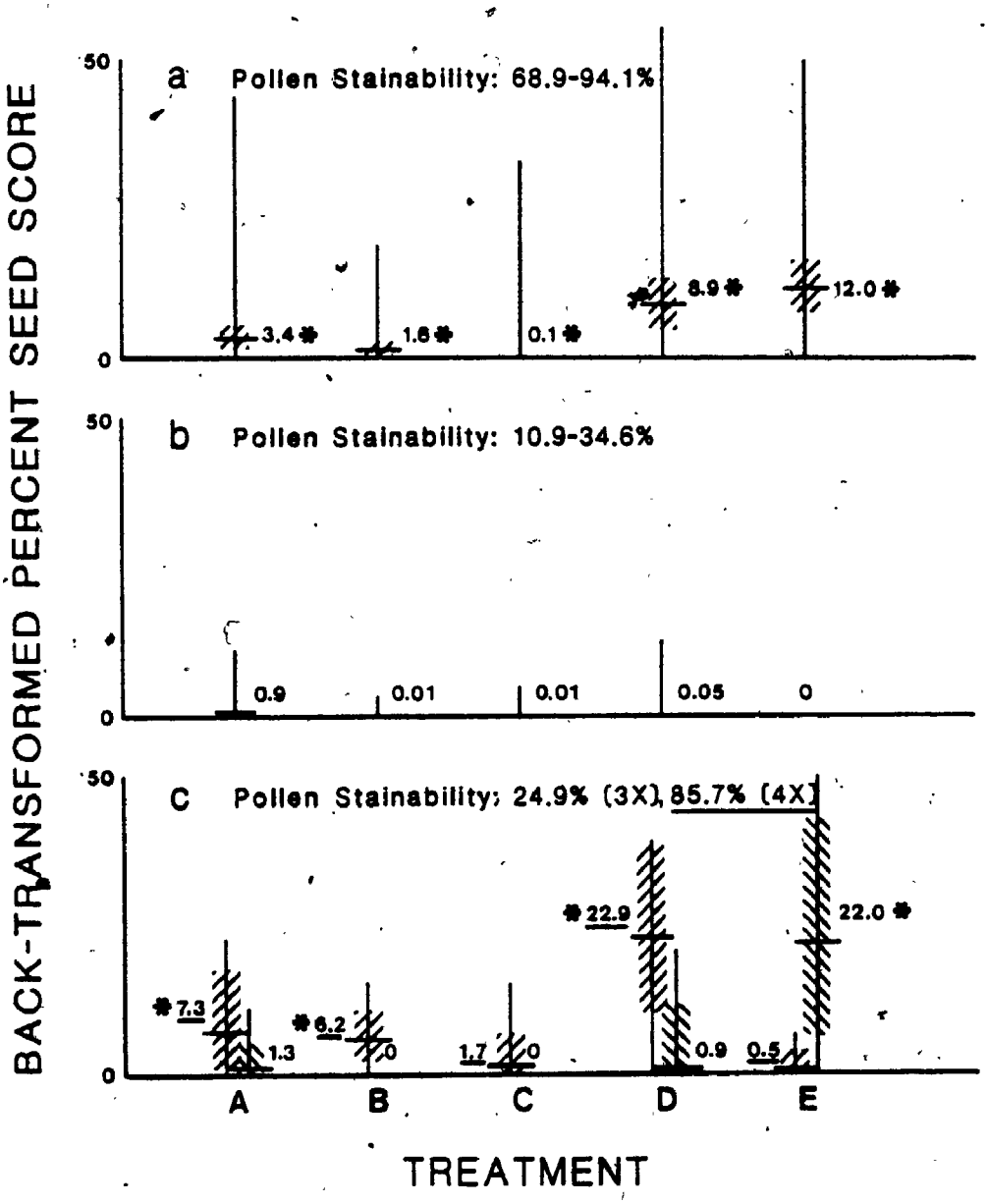
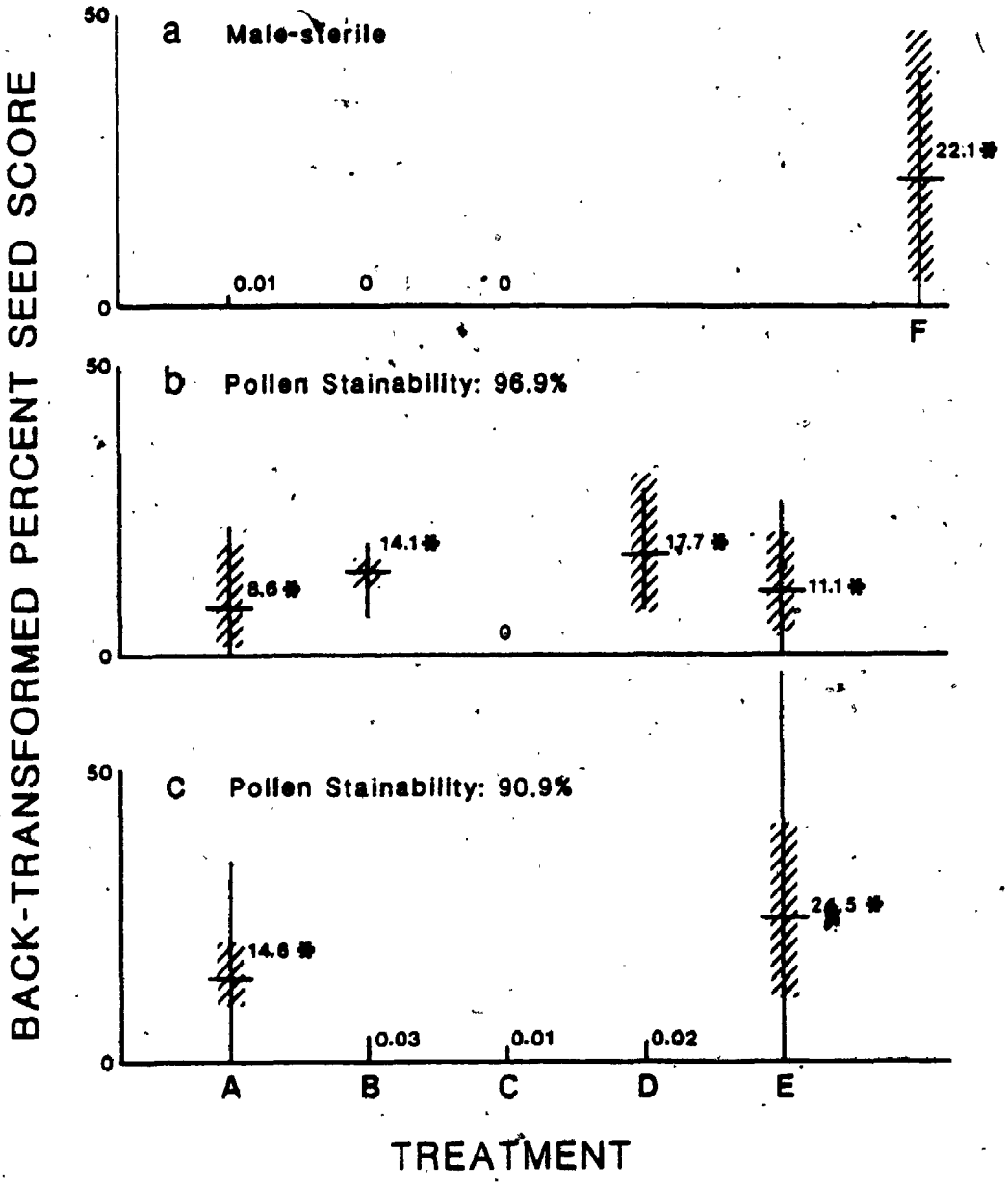

BACK-TRANSFORMED PERCENT SEED SCORE
a Pollen Stainability: 68.9-94.1%
50
0
3.4 *
1.6 *
0.1 *
8.9 *
12.0 *
b Pollen Stainability: 10.9-34.6%
50
0
0.9
0.01
0.01
0.05
0
C Pollen Stainability: 24.9% (3X), 85.7% (4X)
50
0
* 7.3
1.3
* 6.2
0
1.7
0
* 22.9
0.9
0.5
22.0 *
A
B
C
D
E
TREATMENT

Figure 10.3 Experimental seed-set in (a) Crataegus ?grandis; (b) C. sp. aff. C. bushii; and C. punctata. Sample sizes and OTUs as in Table 10.3. Symbols as described for Figure 10.2. Treatments A-F as described in Section 10.5.2. In (a), Treatment F employed pollen from nearby individuals of C. conspecta Sarg. See Appendices 4 and 6 for additional details.


a Male-sterile
22.1 *
0.01
0
0
F
b Pollen Stainability: 96.9%
8.6 *
14.1 *
0
17.7 *
11.1 *
C Pollen Stainability: 90.9%
14.6 *
0.03
0.01
0.02
A
B
C
D
E
TREATMENT
BACK-TRANSFORMED PERCENT SEED SCORE
50
0

Seed-set does not appear to have been limited by stigmatic receptivity under the experimental conditions employed here (Treatments D-F, Fig. 10.2, 10.3). Instead pollination and pollen quality appear to be limiting factors in both C. crus-galli s. l. and C. punctata.

Pollination, even employing pollen from other Crataegus species (Fig. 10.3a) appears to be essential for seed-set to occur. Flowers in which pollination was prevented altogether (Treatment C) had an extremely low rate of seed-set, one that appeared to be correlated with pollen stainability (Fig. 10.2, 10.3). This suggests that what seed-set did occur with this treatment resulted from accidental pollination during or after manipulation.

In 10- and 20-stamen C. crus-galli (Fig. 10.2) and C. sp. aff. C. bushii (Fig. 10.3b) both self- (Treatment D) and cross-pollination (Treatment E) appeared to be equally effective in producing seed-set. However, seed-set with these treatments was also correlated with pollen stainability. Triploid, pollen-infertile 20-stamen C. crus-galli OTUs exhibited very low levels of seed-set with these treatments (Fig. 10.2b) except when highly stainable pollen from a tetraploid 20-stamen OTU was used (Treatment E, Fig. 10.2c).

In C. punctata, however, only open pollination and cross-pollination were capable of producing seed-set (Fig. 10.3c). The self-sterility observed in this species is comparable to that found in other Crataegus species that have been studied to date, and that appear to be exclusively diploid (Bradshaw, 1971, 1975; Love & Feigen, 1978). The self-fertility observed here in three of the morphotypes of C. crus-galli s. l. is reported for the first time in the genus. It may be a consequence of the polyploidy found in these taxa (Table 10.3) since self-sterility in other Maloideae (Pyrus spp.) is of the single-locus gametophytic type (de Nettancourt, 1977) which may be broken down by increases in ploidy (Westwood & Bjornstad, 1971; Grant, 1975).

The lower levels of seed-set with Treatments A and B compared to Treatments D-F in all of the taxa studied here (Fig. 10.2, 10.3) suggest that pollinator service may be limiting seed-set to a considerable degree.

In the morphotypes of C. crus-galli s. l. failure of seed-set with Treatment C (Fig. 10.2; Fig. 10.3a, b) is associated with the presence in the ovules, of the OTUs studied, of multiple embryo-sacs, probably at least in part unreduced ones (Table 10.2). This association tends to confirm suggestions that agamospermy in Crataegus (Muniyamma & Phipps, 1979a) as in the Rosaceae in general (Gustafsson,

1946) is pseudogamous, that is, pollination is required in order for apomictic seed-set to occur.

The incidence of parthenocarpy in fruits developing from emasculated flowers (Treatments C-F) was often higher than in those from other treatments, or in untreated fruits. This was particularly striking with Treatment C (emasculation and isolation only) where although abundant fruit developed in some cases, almost none contained seed (% parthenocarpy 0 - 91.7). This suggests that the results of earlier experiments by Karl Sax, involving emasculation of Crataegus flowers (Section 1.7; Palmer, 1932), may require reinterpretation, since seed-set was apparently never examined.

Seed from experimental pollinations, like that from open pollination, was found to be germinable.

10.6 Discussion

The results obtained in the preceding sections can be interpreted at a number of hierarchical levels, from that of individual trees of C. crus-galli or C. punctata to that of entire topodemes and of taxa. Only the first of these is considered here, leaving the other two levels to the next and final chapter of this thesis. Discussion of the significance of the phenomena described above and in the

appendices for the reproductive success of Crataegus individuals requires reference to be made to a feature of Crataegus ecology described in Section 1.2. This is the way in which many Crataegus species appear to behave as fugitive species, colonizing habitats whose availability is highly unpredictable in space although much less so over time, and persisting there often for only a single generation. In the context of this behavior and the problem of individual reproductive success the phenomena that have been studied here will be discussed in the following five sections under a series of purposely teleological headings that are employed so as to focus attention on the considerations of interest.

10.6.1 Avoiding Hybridization

Given the apparent lack of ecological specialization among Crataegus species indicated by their frequent co-occurrence, at least in Ontario (Fig. 10.4) and in Pennsylvania (Hoover, 1961) it seems reasonable to suppose that offspring of a particular individual can have only a small expectation of becoming established surrounded by conspecific individuals. Under these circumstances it seems likely that selection will favor any change that minimizes the chance of pollinations occurring between species. This is because in the absence of such changes pollen transfers

probably will occur between individuals of congeneric species X and Y. These will result in offspring (assuming fusion of gametes, at the same ploidy level) with only 50% representation of the genotype of either parental individual. If individuals of one species, say X, are a minority locally then with continued cross-pollination the X genotype will come to be represented less and less in successive generations. In terms of ensuring maximum representation of its genotype in future generations an X individual is wasting its gametes and related structures by allowing them to be used by Y individuals, under these circumstances. This will be so even if post-fertilization isolating mechanisms are present, since interspecific pollinations will represent pollen or ovules unavailable for production of exclusively X offspring.

A number of changes may be envisioned which will minimize the chance of interspecific pollinations occurring. One is the change from obligate outcrossing to self-fertility, as seen in C. crus-galli s. l. (Fig. 10.2, 10.3b; Section 10.5.3). Another is to ensure that species X and Y flower only for a short and well-defined period of time. Still another, described by Phipps and Muniyamma (Table 2, 1980) and Smith, Phipps and Dickinson (1980), as well as in Appendix 7 (Fig. A7.1), is the tendency of the individuals of C. crus-galli morphotypes and C. punctata to

flower sequentially rather than simultaneously at any given site where they co-occur.

Control of flowering by means of some mechanism related to vernal heat accumulation (temperature sum; Appendix 7) is a possible means of achieving such a flowering period (Sarvas, 1967). If all individuals of species X at any given site flower when the temperature sum there reaches T_X, the likelihood of pollinations among themselves will be maximized, and that of pollen transfers to and from sympatric individuals of species Y (flowering at T_Y, $T_Y \neq T_X$) will be minimized.

Reference to Figure 10.4 shows that for any of the first six co-occurrence frequency classes (a-f) species pairs are as likely to have significantly overlapping flowering periods (Smith et al, 1980) as not. The only significant exception to this is with the highest frequency class (g), which is represented only by two species pairs whose flowering periods do overlap (C. punctata - C. succulenta and C. monogyna - C. punctata). It is the latter two exceptions that suggest how these data may be interpreted.

Of the 16 species represented in Figure 10.4 C. punctata is the most common. Consequently, it is not surprising that it co-occurs with many other species in the

sample. Its flowering period overlaps significantly with six of them (Smith et al, 1980). Of those with which C. punctata overlaps, Crataegus succulenta and C. compacta appear to be male-sterile (Muniyamma & Phipps, 1979b; Phipps & Muniyamma, 1980) and to produce aposporous embryo-sacs (Muniyamma & Phipps, unpubl.). Given the likelihood that agamospermy in Crataegus is pseudogamous (Section 10.5.3), C. succulenta is unlikely to be under selection pressure against flowering at the same time as a very common species like C. punctata. Instead, C. succulenta probably quite often obtains pollen from C. punctata, with which it frequently co-occurs (Fig. 10.4). Hybrids between sections Macracanthae and Punctatae are not known for Ontario, however, which suggests that either parthenogenetic development of unreduced gametes must be much more frequent than their fertilization, or that some post-fertilization barrier to their establishment exists.

Crataegus monogyna was introduced into North America from Europe probably no earlier than the late 17th century, and has since become established here (Palmer, 1963; Phipps & Muniyamma, 1980). The result of the failure of C. monogyna in North America to flower at a time different from that at which common native species do is seen in the occurrence of C. monogyna hybrids, with C. punctata in Ontario (Phipps & Muniyamma, 1980) and with C. douglasii

var. suksdorfii in Oregon (Love & Feigen, 1978).

The occurrence of these hybrids is probably also a consequence of the way in which all three of these taxa, as well as C. laevigata (Poir.) DC (with which C. monogyna hybridizes in Europe; Bradshaw, 1971, 1975) are self-sterile. Their seed-set depends on receiving pollen from other individuals. With the breakdown of geographic barriers (and in Europe, ecological ones; Bradshaw, 1971; Byatt, 1975, 1976a,b), these other individuals flowering at the same time may belong to other congeneric species.

In the case of C. crus-galli s. l. data is available only for the 10- and 20-stamen morphotypes (Fig. 10.4; Appendix 7, Fig. A7.1). The flowering periods of these two overlap with each other, and with those of C. monogyna, C. punctata and C. succulenta, at the level of resolution shown in Figure 10.4 (Smith et al, 1980). Comparison with Figure A7.1 suggests that the overlap in the flowering periods of the two crus-galli morphotypes and C. punctata is probably more probabilistic than consistent from year to year (Appendix 7). In addition, in contrast to the self-sterile taxa discussed above, the morphotypes of C. crus-galli, particularly the highly pollen-fertile ones, are protected from the effects of interspecific pollen transfers to some degree by their self-fertility, if not also by their apomixis as well. The extent to which this is

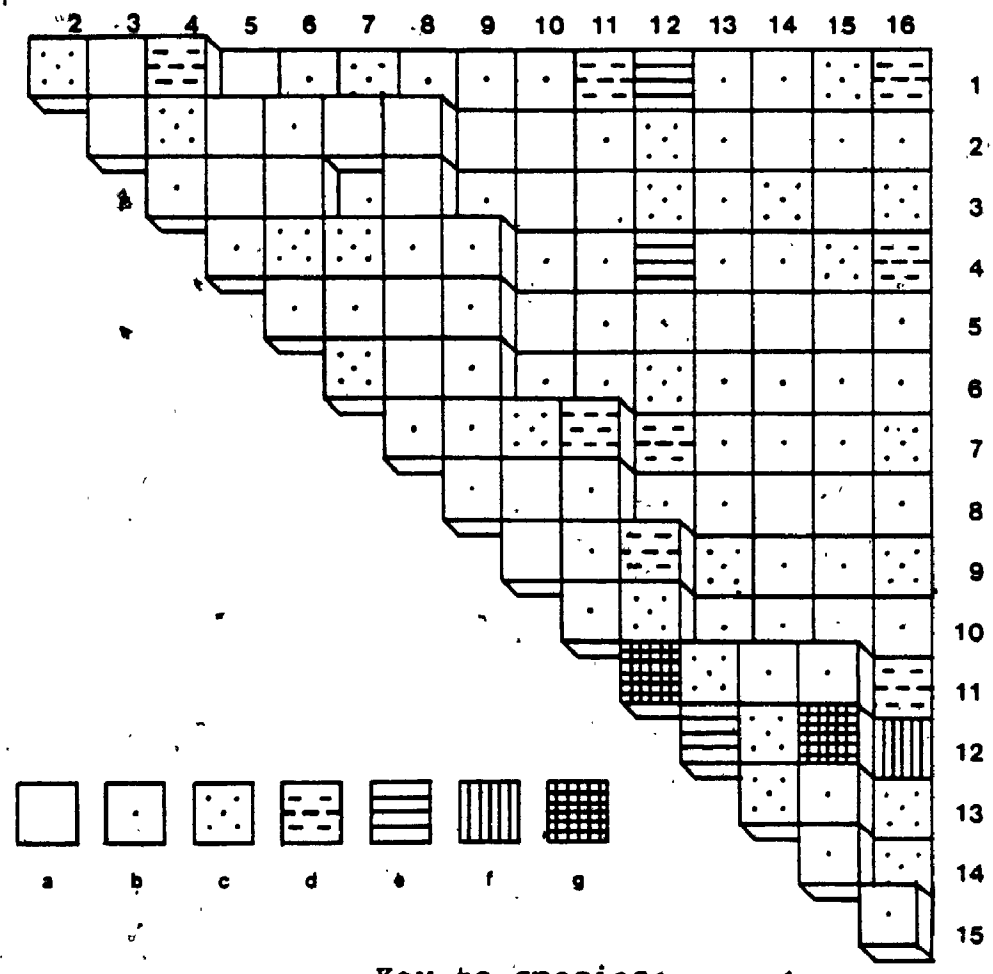
true is suggested by the fact that while individuals morphologically intermediate between C. section Crus-galli and section Punctatae are found (Palmer, 1946), they are quite infrequent (Section 11.5).

The overlap of the flowering periods of male-sterile C. succulenta and 10-stamen C. crus-galli (Fig. 10.4) probably represents the same phenomenon as that with C. succulenta and C. punctata: lack of selection against flowering at the same time as a common and often abundant pollen-fertile species. However, as described below (Section 11.5.3.1), the occurrence of hybridization between section Crus-galli and section Macracanthae has been suggested (Eggleston, 1908a, 1923), and may be responsible for the morphotype referred to here as C. sp. aff. C. bushii.

Hybridization between C. crus-galli s. l. and C. monogyna has not been encountered. It is unclear how hybridization between 10- and 20-stamen C. crus-galli could be recognized morphologically.

The fairly consistent relative flowering times seen at Sites 1 - 5 for 10- and 20-stamen C. crus-galli and C. punctata (Fig. A7.1) thus probably represent the results of the successful transmission of the genotypes of ancesteral individuals whose flowering periods overlapped least with those of sympatric taxa.

Figure 10.4 Data for 728 Ontario collections of Crataegus species (1-16). Shading (a-h) indicates number of sites (out of a total of 391) at which species occur together: (a) 0-3; (b) 4-7; (c) 8-11; (d) 12-15; (e) 16-19; (f) 20-23; and (g) 24-27 (P. F. Ulf-Hansen, personal communication). Elevated portion of the semi-matrix along the diagonal indicates species pairs for which the 95% confidence intervals of each species' mid-anthesis temperature sum overlap (P. G. Smith, personal communication; Smith, Phipps, & Dickinson, 1980). Sampling for this compilation was not random, but rather was based (mid-anthesis) on herbarium material at UWO, and (co-occurrence) on all Ontario material examined and mapped by Phipps and Muniyamma (1980). Co-occurrence at a site means that recorded collection localities differed by less than 0.01 degrees latitude and longitude.



Key to species:

(1) C. pedicellata Sarg.; (2) C. holmesiana Ashe; (3) C. mollis (T. & G.) Scheele; (4) C. pringlei Sarg.; (5) C. schuettei Ashe var. schuettei; (6) C. macrosperma Ashe var. macrosperma; (7) C. flavida Sarg.; (8) C. suborbiculata Sarg.; (9) C. compacta Sarg.; (10) C. pruinosa var. cognata (Sarg.) Phipps; (11) C. succulenta Lindl.; (12) C. punctata Jacq.; (13) 20-stamen C. crus-galli L.; (14) 10-stamen C. crus-galli L.; (15) C. monogyna Jacq.; and (16) C. calpodendron (Ehrh.) Medic.

10.6.2 Ensuring Pollination

The results of Section 10.2 (see also Appendix 4) concerning pollen production in self-fertile C. crus-galli s. l. can be seen as the result of a lack of other floral specializations to ensure pollinator service (Appendix 3; Section 10.4), coupled with a requirement for pollination in order for seed to be set whether sexually or otherwise (Section 10.5).

The high P/Os found in the 10- and 20-stamen C. crus-galli and C. sp. aff. C. bushii (Fig. 10.1) are comparable to those associated in C. punctata (Fig. 10.1; Fig. 10.3c) and other taxa (Cruden, 1977) with either facultative or obligate xenogamy. High P/Os are likely to have been maintained in the crus-galli morphotypes because of the dependence of both sexual and apomictic seed production on the occurrence of pollination. Anther dehiscence in unopened flowers is infrequent, although it has been observed occasionally. Once flowers have opened, opportunities for pollination (autogamous, geitonogamous or xenogamous) appear to be limited to those provided by insect visitors, notably pollen-collecting bees (Section 10.4). Obtaining pollinator service may thus depend on continuing high levels of pollen production (compare Treatment A, Fig. 10.3a, with that in Fig. 10.2, 10.3b, c).

Cruden and Miller-Ward (1981) have also refined earlier interpretations of P/Os. They argue that these ratios should be inversely proportional to the likelihood of a pollen grain reaching a stigma, which is the ratio of stigmatic area to the pollen-bearing area of the pollinator. The stigmatic areas of C. crus-galli s. l. are small (Appendix 3, Plate 8b; approximately 0.25 mm per stigma) compared to the size of the predominant pollinators. In addition, in some individuals effective (stainable) pollen grains may be diluted by large numbers of ineffective non-stainable pollen grains. These, as well as incompatible pollen grains, may also tend to clog the limited stigmatic surfaces. Dependence on pollen-gathering pollinators together with pseudogamy and in some cases pollen infertility, as well as small stigmatic areas all probably contribute to selection for high levels of pollen production, offsetting the effect seen in other taxa of self-fertility.

10.6.3 Maximizing Investment Returns - Apomixis

In their paper on the evolution of apomixis Marshall and Brown (1981) review three alternative explanations of the role of apomixis in higher plants. Two of these are of interest here, while the third (the "Model T hypothesis") will be discussed in Chapter 11.

The oldest current explanation of apomixis is that it represents an "escape from sterility" (Darlington, 1939; Stebbins, 1941, 1950) on the part of otherwise sterile hybrids. Certainly in the case of the triploid OTUs found here (Chapter 9) among 20-stamen C. crus-galli and in C. ?grandis apomixis probably does represent the means by which these OTUs are able to produce seed (Appendix 3, Fig. A3.2). Successful meiosis is unlikely in these OTUs, although a situation like that in the Rosa canina complex (Stebbins, 1950) cannot be ruled out.

As noted above (Section 10.6.1) cross-pollination and fertilization result in only 50% representation of each parent's genotype in the offspring (assuming as before fusion of gametes of the same ploidy level). Even pollen exchange with conspecific individuals will have the same result, at least in the short-term, and assuming parents are unrelated. Mutations enabling uniparental production of offspring, whether by means of apomixis or autogamy will eventually become fixed in plant populations unless at the same time they drastically reduce fitness in some other way (Marshall & Brown, 1981; Jain & Workman, 1967). This is the "automatic advantage" hypothesis of Williams (1975), Maynard Smith (1978) and others (cited by Marshall & Brown, 1981).

Such uniparental monopolization of offspring will, if complete, result in limited opportunities for the introduction of evolutionary novelty into any such line. However, if opportunities for gamete fusion remain then genetic variation will be maintained within a line (Marshall & Weir, 1979). Given gametophytic apomixis such as the apospory found in Crataegus, (Section 10.3; Muniyamma & Phipps, 1979a and unpubl.), then pseudogamy (Section 10.5.3) practically ensures that some level of crossing will persist. In the tetraploid OTUs studied here (10- and 20-stamen C. crus-galli; C. sp. aff. C. bushii) facultative sexuality may also be possible (Section 10.5.3; Table 10.2), further ensuring maintenance of genetic variation within lines of descent.

Another aspect of the role of apomixis in increasing the likelihood of seed-set, particularly in pollen-infertile triploids, may be the role of aposporous embryo-sacs in extending the longevity of ovules, thus increasing their "effective pollination period," as described for apples by Williams (1970).

10.6.4 Minimizing Pollen and Ovule Wastage - Self-fertility

As discussed elsewhere (Section 10.5.3; Chapter 11) the self-sterility observed in diploid Crataegus taxa (Fig. 10.3c; Bradshaw, 1971, 1976; Love & Feigen, 1978) may be

broken down by increases in ploidy such as may result from the fertilization of unreduced gametes (Chapter 11; Camp & Gilly, 1943). In the triploid and tetraploid morphotypes of C. crus-galli s. l. studied here self-pollinations appear to be equally as effective as cross-pollinations in producing seed, assuming the availability of fertile pollen (Fig. 10.2a, c; Fig. 10.3b). The individuals of these taxa thus appear to be maximizing representation of their genotypes in their offspring, and also minimizing the limitation imposed on seed production by pseudogamy, as well as minimizing the extent to which pollen is wasted by transfer to incompatible stigmas. However, such minimization does not appear to have resulted in an opportunity to invest less in pollen production (Section 10.6.2).

10.6.5 Moving On - Dispersal and Establishment

Reference was made above (Section 10.6) to the way in which hawthorn genotypes -in eastern North America at any rate- apparently have to rely on moving successive generations from one site to another. Movement of successive generations requires dispersal; but as suggested in Appendix 5 this aspect of hawthorn biology is poorly understood at present, morphological adaptations for endozoochory notwithstanding.

Similarly, little is actually known about how Crataegus topodemes become established, either in terms of comparisons among species with respect to the relationship between their seed output and their representation in developing stands, or with respect to changes in stand composition with time.

10.6.6 Conclusions

This chapter has described a suite of characteristics of reproduction in C. crus-galli s. l., notably self-fertility, agamospermy, pseudogamy and high pollen-ovule ratios, together with considerable variation in pollen fertility. In and of themselves these are not at all novel features in flowering plants. Taken together, these features represent one possible, coherent pattern of reproductive function in the case of the pollen-fertile tetraploid morphotypes. A different but equally coherent pattern is exhibited by the self-sterility, sexual seed production and high pollen-ovule ratios of diploid C. punctata. In the case of the triploids, notably the pollen-infertile 20-stamen C. crus-galli it would appear that reproduction in these individuals may be limited by the availability of fertile pollen from individuals of other Crataegus taxa (Fig. 10.5b, c). The same limitations probably apply even more stringently to the male-sterile OTUs of C. ?grandis at Site 4 (Fig. 10.4, 10.6a).

The real import of these patterns of reproductive function documented in Crataegus for the first time here lies in the consequences they may have for the degree of morphological variability to be observed within and among topodemes (Chapter 7; Section 8.7). This is the key to understanding not only evolution within Crataegus species groups but also past perceptions of the results of that evolution, namely earlier Crataegus taxonomies. These considerations are discussed in the next chapter, leading to the conclusion that future taxonomic treatments of groups such as C. crus-galli s. l. must take into account both the kinds of variation and variability patterns described in earlier chapters and the reproductive characteristics discussed here to which these patterns probably owe their origin.

CHAPTER ELEVEN

GENERAL DISCUSSION

11.1 Introduction

In order to provide insights into the nature and origins of the patterns of morphological variation in Crataegus this thesis has presented quantitative descriptions of those patterns in C. crus-galli s. l. and C. punctata with reference to a suite of reproductive and vegetative descriptors. These results are reviewed in the first part of the following section (Section 11.2.1).

These descriptors also document significant variation among the taxa studied in the degree of variability associated with these patterns. Data obtained on reproductive behavior and variation in ploidy level are suggestive of the events and processes probably responsible for observed patterns of morphological variation and differences in variability. These are reviewed in Section 11.2.2.

Section 11.3 presents a model of evolution in C. crus-galli s. l. which makes the connection between these two sets of data by reference to existing knowledge of

Crataegus ecology and models of evolution in this genus and others like it.

The significance of this model for Crataegus taxonomy is then discussed in Section 11.4. The chapter concludes with a commentary on the taxonomic treatment of Ontario C. crus-galli s. l. in light of the conclusions that have been reached here (Section 11.5).

11.2 The Significance of Topodeme Structure

Topodeme structure may be described with reference to a suite of morphological descriptors (Table 3.1, 3.3) by the location of a topodeme sample mean vector, and by the pattern and magnitude of the dispersion of such a sample around its mean vector. The results obtained concerning the structure of topodemes of C. crus-galli s. l. (T1 - T6) and the topodeme of C. punctata (T7) are summarized below.

11.2.1 Topodeme Structure in C. crus-galli s. l.

The individual topodeme samples of C. crus-galli s. l. (T1 - T6) and of C. punctata (T7) were consistently recovered in the cluster analyses of the entire 160 OTU sample (Chapter 5; Section 5.3) based on all eleven reproductive descriptors (Table 3.1). Similar results were obtained also with the 60 OTU subsample in analyses based on

all descriptors available or on exclusively reproductive or exclusively multistate ones (Section 6.3; Table 4.2). Leaf data available for the 60 OTU subsample led to less complete recovery of the topodeme samples (Section 6.3). This was due to their greater variability with respect to these descriptors (Table 7.4). These analyses also demonstrated that much of the structure of the samples examined was independent of patterns of descriptor correlation (Fig. 5.7, 5.11; Fig. 6.13; Section 6.4).

Differentiation of the topodeme samples was also observed in ordinations: Principal components analyses (Section 6.2) and canonical correlation analyses (Fig. 8.1; Section 8.4.1) all demonstrated the distinctness of the topodeme samples.

Analyses testing explicit a priori groupings of OTUs (canonical variates analyses, Section 8.5.1, 8.5.2; principal coordinates analyses, Section 8.6; R-PCA of the pooled within-groups variance-covariance matrix **W**, Section 8.7; and comparison of leaf spectra, Section 8.8), confirmed the differentiation of the seven topodeme samples.

Such ordinations were also useful because they provided low-dimensional summaries of descriptor variation among topodeme samples. These demonstrated the way in which the four topodeme samples of 10-stamen C. crus-galli (T2, T3,

T5, T6) are distinguished from the two 20-stamen C. crus-galli ones (T1, T4) by descriptors other than stamen number. Similarly, descriptors other than pubescence were shown to distinguish the punctata topodeme sample (T7) from the crus-galli ones (Fig. 8.3, 8.4; Table 8.8d, 8.9d, 8.11e).

Comparison of topodeme sample covariance matrices for six flower and fruit descriptors indicated that except possibly for T6, all seven topodeme samples were little differentiated from one another (Section 7.6). Comparison of the individual topodeme sample covariance matrices with the pooled within-groups covariance matrix $\underset{\sim}{W}$ for the 93 OTU sample of 10- and 20-stamen C. crus-galli (Section 8.9.3.2; Table 8.18) indicated a somewhat greater degree of differentiation, notably of both T6 and T7 (C. punctata) from the remainder of the crus-galli samples. The former set of results suggests that the descriptors used relate to only a limited portion of the phenotype, within which their covariation is largely constrained to a common pattern regardless of taxon.

While overall patterns of reproductive descriptor covariation are similar in all seven topodeme samples, their variability with respect to the flower and fruit descriptors differs markedly between the crus-galli topodeme samples (T1 - T6) and that of C. punctata (T7; Section 7.5; Table

7.3, 7.4). This differentiation is observable in all of the ordination spaces examined (Fig. 6.1a, 6.2a, 6.4-6.7, 8.3, 8.4), as well as in the cluster analyses (Section 5.4).

These results suggest that the complexity of morphological variation in groups such as C. crus-galli s. l. is a function of many relatively small but correlated differences in individual descriptors between topodemes, combined with quite low levels of variability in these descriptors. While diligent univariate analyses have in the past provided tantalizing evidence of such patterns of covariation (Rickett, 1936, 1937), the use here of a wide range of multivariate methods has proved much more effective. A number of these methods are also extremely valuable because of the description they provide of the overall variability of the samples examined.

11.2.2 Topodeme Structure as a Probable Consequence of Reproductive Behavior

In terms of reproductive behavior the contrast seen here between C. crus-galli s. l. and C. punctata is that both apomixis and self-fertility are present in the former, while absent in the latter. All four Ontario morphotypes of C. crus-galli s. l. produce apparently unreduced, aposporous embryo-sacs (Section 10.3). Results of pollination experiments (Section 10.5) showed that the three

more or less pollen-fertile morphotypes (10- and 20-stamen C. crus-galli, C. sp. aff. C. bushii; Section 10.2) are self-compatible, and that in all four agamospermy is pseudogamous.

As described in Section 10.6.3, agamospermy represents a means by which almost any genotype that arises may be enabled to persist for more than a single generation, regardless of its sexual fertility. As pointed out by Clausen (1954) this means on the one hand that sexually sterile genotypes are able to reproduce (the "escape from sterility" hypothesis, Section 10.6.3). On the other hand, with gametophytic apomixis (apospory, diplospory) where sexual processes may continue to occur, however infrequently, the effect is only to slow down the tempo of genotypic change through recombination. The same is true of self-fertility. Nor are these two effects mutually exclusive. Employing the analogy of the automobile assembly line, Clausen (1954) advanced the "Model T hypothesis" (Marshall & Brown, 1981):

> "Once a model of an apomict has been developed, innumerable replicas of it can be produced. If it is a successful one, the replicated model becomes very common until the requirements change and it no more is fitted to its task. Then new models evolved through more or less radical interchanges and alterations get their chance, and swarms of replicas of them are turned out. Not all the old models disappear simultaneously, but old and new travel together, although the new may outrun the

> old. ... Facultative apomixis, therefore, does not prevent variation; rather, it multiplies certain varietal products." (Clausen, 1954)

Because the "requirements" of the environment encountered by a colonizing species may change unpredictably, rapid genetic adaptation to a current habitat may result in features that are disadvantageous in a new one (Levin, 1975). In a relatively long-lived polycarpic perennial like Crataegus the genetic conservatism of apomixis and self-fertility may be advantageous not only for this reason but also because of the relatively large number of progeny that may be produced by each individual during its life-span.

The topodeme structure observed here in C. crus-galli s. l. appears to result from the interaction between features of its breeding system described above and in Chapter 10, and other characteristics of reproduction in the genus as a whole. Among the latter are the discreteness of the sites at which seedling establishment is possible, and the way in which Crataegus seed deposition is probably clumped as a result of dispersal by feeding animals with non-random patterns of movement. Given apomixis and self-fertility, the rain of crus-galli seed on a given site is likely to have resulted from a very small number of different reproductive events involving only one or a few

parent individuals.

The self-fertility and potential for agamospermy observed in the crus-galli topodemes (T1 - T6) are seen not as the cause of the topodeme uniformity also observed, but rather as being indicative of the circumstances under which the topodeme in question was brought into existence.

In this way, a crus-galli topodeme is seen as the result of a large number of predominantly sibling, predominantly uniparental individuals arriving at a site over a relatively short period of time. Monopolization of a site by these individuals will be enhanced if they also succeed in establishing their own offspring there as well.

Differences among topodemes arise out of the potential uniqueness of the reproductive events by which each one is established. These differences are also a function of both the unpredictability of the distance and direction of seed dispersal, and the availability of sites suitable for seedling establishment.

In the case of the topodeme of C. punctata (T7), its greater morphological variability can likewise be seen as the result of antecedent reproductive events. Again, the nature of these events is indicated by the reproductive behavior observed. Given self-sterility and the absence of apomixis the individuals present must be the offspring of at

least two different individuals. Under the same general set of conditions as described above for C. crus-galli s. l. (clumped dispersal, colonization by sibling individuals) there are also advantages accruing to continued sexuality and self-sterility, despite the potential disadvantages discussed earlier (Section 10.6). Disadvantages of sibling competition are minimized with increasing genetic variance among siblings brought about by sexual outcrossing (Smith, 1978). Moreover, under conditions of intense selection at a new site very successful genotypes are more likely to be found among sexually produced progenies. If the environmental conditions encountered by successive generations vary unpredictably (as a result of relatively long-distance dispersal or, with short dispersal distances, as a result of the presence of parent individuals) then the breakdown of highly successful parental genotypes ("sisyphean genotypes," Williams, 1975) by sexual reproduction will not be disadvantageous (Williams, 1975; Smith, 1978).

11.3 A Model for Evolution in Crataegus crus-galli L. sensu lato

By adapting a description of agamic complexes in Crepis given by Babcock and Stebbins (1938), Camp (1942) and Camp and Gilly (1943) proposed a model for the origin and

maintenance of polyploid species complexes in Crataegus. According to this model the occurrence and occasional fertilization of unreduced gametes in a limited number of ancestral diploid species, together with hybridization, was thought to have been responsible for the wide range of frequently polyploid forms exhibited by the genus in North America (Camp, 1942; Camp & Gilly, 1943). Elaboration of their model in light of the results obtained here, and with reference to C. crus-galli s. l. follows.

11.3.1 Assumptions of the Model

The model developed here assumes that the locus of diversification of C. crus-galli s. l. corresponded approximately to the modern range of the complex. To the extent that the most recent diversification of this complex occurred during, or at the end of the Pleistocene glaciation, this range should be shifted south beyond the glacial margin, and perhaps also east onto parts of the continental shelf exposed by lowered sea levels as well as southwest toward Mexico. In this case the processes described below would also have been affected by post-Pleistocene migrations northward of ancestral species (Sinnott & Phipps, 1983).

Ancestors of C. crus-galli s. l. are assumed to have been self-sterile diploids in which formation and

parthenogenetic development of unreduced gametes were absent. The habitat requirements of these ancestors, however, are assumed to have been little different from those exhibited today by most of the genus as a whole (Section 1.2). Such ancestral forms would have been similar to modern species like C. punctata in these respects.

The abundance of these crus-galli ancestors is assumed to have been limited by the availability of suitable sites for establishment, such as on floodplains and erosion surfaces (Marie-Victorin, 1938; Whitney, 1982). Prior to Amerindian and then European land-clearing for agricultural purposes forest margins may not have been sufficiently frequent to have provided significant opportunities for hawthorn establishment.

Under these conditions, and with a lack of ecological specialization among most Crataegus species (Fig. 10.4; Section 10.6.1), selection for flowering periods narrowly defined in terms of a criterion like vernal heat accumulation (Appendix 7; Section 10.6.1) is assumed to have already occurred in the crus-galli ancestors and sympatric, congeneric species.

11.3.2 Acquisition of Apomixis

Given their multicellular archesporium (Section 10.3; Davis, 1966) it may be that the apospory frequent in several rosaceous genera (Gustafsson, 1947b) resulted from small changes in the control of development in archesporial cells, such that gametophyte development no longer required prior meiosis. If a second mutation enabling parthenogenetic development of gametes also occurred (Marshall & Brown, 1981) then aposporous apomixis would have been achieved in these genera.

The occurrence of such mutations in crus-galli ancestors would have had a dramatic effect on their reproductive success. Initially, even with pseudogamy, apomictic genotypes would have become represented by more and more individuals, since any compatible pollination would induce seed-set without affecting progeny genotype (see Section 11.2.2 above).

11.3.3 Polyploidy and Self-fertility

Variation in ploidy level in ancestral C. crus-galli could have come about through the occasional fertilization of unreduced gametes (Camp, 1942; Camp & Gilly, 1943; Harlan & deWet, 1975). Repeated crosses involving reduced and unreduced gametes would readily have resulted in the

occurrence of triploids and tetraploids in addition to diploids. Given the nature of self-incompatibility in the Maloideae (Section 10.5; de Nettancourt, 1977) it is possible that such increases in ploidy would have conditioned self-fertility at the same time. While triploids would have been disadvantaged by their consequent pollen-infertility (assuming unreduced microspores could not be produced), tetraploids would have gained the double advantage of being enabled to reproduce sexually as well as apomictically (Nygren, 1967), as well as of ceasing to be dependent on the availability of other individuals for compatible pollinations.

The infrequency of polyploids above the tetraploid level observed here in Ontario C. crus-galli s. l. (Chapter 9) may result from a combination of mechanisms, including ones similar to those apparently operating in the genus Rubus. These include phenological separation not only of most species but also of triploid and tetraploid crus-galli morphotypes (Appendix 7), and the apparent absence of diploid crus-galli individuals (Chapter 9). In Rubus crosses between facultatively apomictic polyploids tend to result in predominantly maternal offspring. Predominantly biparental offspring are produced only when there is a marked disparity in ploidy between the parents, notably when pollen from diploid individuals is transferred to the

stigmas of a polyploid (Sax, 1954).

11.3.4 Hybridization

It may be argued that the diploid, self-sterile Crataegus taxa seen today (or in the recent past) as distinct entities (e.g. C. punctata, C. monogyna, C. laevigata; Muniyamma & Phipps, 1979b; Bradshaw, 1975) are seen as such because they are represented by individuals whose ancestors successfully avoided hybridization by ecological or phenological specialization or other means.

The occurrence of apomixis, however, enabled almost any hybrid that did arise to persist unaltered for generations regardless of its sexual fertility. Given pseudogamy, self-fertility and even partial pollen-fertility like that found in triploid 20-stamen C. crus-galli may be sufficient to ensure at least a low level of seed production (Fig. 10.2c), and hence persistence. Given persistence of the F_1 genotype in this way, fertilization of its unreduced gametes would undoubtedly also occur eventually.

Taxa represented today by polyploid individuals may or may not have been as successful in avoiding hybridization as the diploid ones referred to above. Some modern Crataegus taxa represented exclusively by triploids or tetraploids may be examples of the latter case. If no recognizably related

diploids exist then it may be that such polyploids arose through hybridization. Alternatively, if the diploid count obtained by Longley (1924) for C. canbyi (range: Pennsylvania, Delaware and Maryland; Palmer, 1925) is indicative of the persistence south of the glacial margin of diploid C. crus-galli s. l. then this species complex may be an example of the former. The occurrence of all three ploidy levels within C. crus-galli s. l., coupled with its morphological distinctiveness (Palmer, 1946; Phipps & Muniyamma, 1980) would suggest that here apomixis, autopolyploidy and self-fertility may have helped to largely isolate the complex from contamination by other species groups.

The occurrence throughout most of the range of the complex of both 10- and 20-stamen morphotypes (Palmer, 1925; Rickett, 1937) is open to a number of possible interpretations. Of the two, 10-stamen C. crus-galli is the more widespread and may more closely resemble the ancestral form if it can be shown to include the diploid individuals (C. canbyi is a 10-stamen form of C. crus-galli s. l.; Palmer, 1925). If this is so then the 20-stamen morphotype could represent the results of an old hybridization and consequent introduction of the 20-stamen trait into C. crus-galli, one that has spread and persisted in southern Ontario despite triploidy and pollen-infertility (Table 9.2;

Section 10.2). Alternatively the 20-stamen morphotype may be a hybrid of polytopic and much more recent origin. These possibilities are discussed further in Section 11.5.3 below.

11.3.5 The Effect of Human Disturbance

According to the model being developed, prior to European agricultural activity in eastern North America hawthorns like C. crus-galli s. l. were restricted largely to naturally occurring open, frequently disturbed habitats (Marie-Victorin, 1938; Palmer, 1946; Whitney, 1982). When because of extrinsic reasons land formerly cleared and used agriculturally was abandoned, it was readily occupied by species requiring high light intensities and adapted to colonization of disturbed soils. In New England extensive stands of white pine developed on abandoned farmland during the early part of the nineteenth century (Raup, 1966), as they apparently did elsewhere, on the sites of Amerindian cultivations. On the heavier soils found west of the Appalachians recolonization probably more frequently involved hawthorns.

As a result of the increased availability of suitable sites during the latter part of the nineteenth century large hawthorn colonies developed throughout eastern North America (Sargent, 1907; Brainerd, in Brown, 1910; Palmer, 1946). Possibly because of their larger size and increased

frequency, colonization of these sites resulted in very mixed assemblages of taxa (Sargent, 1907), with the further result that the opportunity for hybridization increased dramatically (Brainerd, in Brown, 1910; Marie-Victorin, 1938). Thus the very large numbers of Crataegus entities recorded in North America during the latter part of the nineteenth, and early twentieth centuries appears to have resulted from the interaction between the reproductive behavior and ecology of the genus, and the relatively sudden increase in the availability of of suitable habitats, quite apart from any peculiarities of taxonomic philosophy on the part of the workers describing them.

In a model of Crataegus evolution similar to the one presented here, Sinnott and Phipps (1983) have pointed out that the events in North America of the last five hundred years, and their effects on genera like Crataegus, are in fact superimposed not only on Pleistocene and post-Pleistocene events but also on an entire sequence of quaternary glacial-interglacial cycles (Davis, 1976). Sinnott and Phipps (1983) suggest that peribds of speciation were associated with the glaciations. Conversely, episodes of intense selection in the course of the reestablishment of forest cover during the interglacial periods acted to stabilize the genus (Sinnott & Phipps, 1983).

11.3.6 Comparison with Other Genera

The model proposed above on the basis principally of the results obtained in Chapters 9 and 10, together with the ideas of Brainerd, Brown, Palmer, Marie-Victorin and Camp can be compared with similar models for other species complexes affected by the occurrence of apomixis, polyploidy and hybridization.

Within the Maloideae three other genera are notoriously complex, taxonomically. These are Aronia, Amelanchier and Sorbus. Two distinct eastern North American species of Aronia Medic., A. arbutifolia (L.) Ell. (2n = 34, 68) and A. melanocarpa (Michx.) Ell. (2n = 34) apparently hybridize to produce an intermediate form, A. prunifolia (Marsh.) Rehder (Robertson, 1974). Hardin (1973) presented evidence that all three species are self-fertile, and that A. prunifolia reproduces apomictically. However, in the latter case the occurrence of agamospermy and parthenocarpy were not distinguished from each other. Variation in reproductive behavior evidently occurs since Nova Scotian A. melanocarpa have been shown to be self-sterile (Hall, Wood & Jackson, 1978).

In the Rosaceae as a whole the taxonomic complexities of the genus Amelanchier may be exceeded only by those of Crataegus and Rubus (Robertson, 1974). As in Crataegus

diploid, triploid and tetraploid species are known (Robertson, 1974; Robinson & Partanen, 1980). Some taxonomists dismiss the occurrence of hybridization in the genus (Jones, 1946), as well as that of apomixis (Robertson, 1974; Robinson, 1982). As yet neither conclusive experimental nor anatomical evidence is available concerning apomixis; the occurrence of polyploidy, however, is suggestive.

The genus Sorbus L. in Europe has been studied in much greater detail than have the two preceding genera (Liljefors, 1953, 1954; references cited by Richards, 1975, and Challice & Kovanda, 1978). It is represented there by three diploid species, S. aria (L.) Cr., S. aucuparia L. and S. torminalis (L.) Cr. Post-Pleistocene migration of tetraploid S. aria into Scandinavia accompanied by (diploid) S. aucuparia and S. torminalis apparently resulted in a number of crosses and back-crosses between S. aria and the other two species to produce a range of triploid and tetraploid agamospecies (Liljefors, 1954). Elsewhere diploid hybrids that are moderately pollen-fertile have been produced by crosses between S. aria and S. aucuparia (Richards, 1975). The relationships of the species and agamospecies of Scandinavian Sorbus are probably simplified by the limited number of sexual species involved, as well as possibly also by their relatively recent migration into the

area they occupy. The situation in central Europe (Kovanda, 1961) and in eastern Asia (Stebbins, 1950) appears to be more complex. In the latter case, as with Crataegus in North America, human disturbance appears to have also played a part (Stebbins, 1950).

The model described above for evolution in C. crus-galli s. l. and the one discussed by Sinnott and Phipps (1983) have been developed not only from results obtained with Crataegus and other Maloideae but also from those of studies of other rosaceous genera (e.g. Alchemilla, Gustafsson, 1947b; Rubus, Fernald, 1950; Haskell, 1966; Potentilla, Asker, 1979) and of genera in other families, notably the Asteraceae and Poaceae (reviewed by Gustafsson, 1946, 1947a, b; Nygren, 1967; Asker, 1979). Studies of the genetics and evolutionary history of these genera, in which apomixis, hybridization and attendant polyploidization have also resulted in large complexes of taxa, have revealed situations broadly similar to those hypothesized here. With North American Crataegus, however, much more work is required in order to document actual phenomena. At present C. crus-galli s. l. is one of the best known groups within the genus. This knowledge is based only on the present morphometric and reproductive studies in Ontario, the morphometric studies of Rickett (1937) in Missouri, and on merely thirty chromosome counts from

Ontario alone (Table 9.2; Muniyamma & Phipps, 1979b). Comparable morphometric data and documentation of chromosome numbers from the remainder of the range of this complex are the minimum required in order to develop a more complete picture of its evolutionary history.

11.4 Implications of the Model for *Crataegus* Taxonomy

The implications for *Crataegus* taxonomy of topodeme structure and its probable origin in the reproductive behavior of *Crataegus* taxa are both historical and current. Historically they assist in understanding how "The *Crataegus* Problem" (Palmer, 1932; Camp, 1942) arose. They also suggest how it may be advisable for modern *Crataegus* taxonomic research to proceed, methodologically and in terms of the kinds of taxon concepts and classifications that may be useful.

Palmer (1946) reviewed the factors which, in his eyes, accounted for the enormous number of species described for North American *Crataegus*, particularly during the period 1892-1925. He echoed Provencher (1862; Section 1.1) in referring to the "undue importance given to slight variations in foliage, flowers, and fruit." Like some of Brown's respondents (Brown, 1910; Eggleston, 1910), he suggested that the earlier students of the genus often worked too independently of each other, simultaneously, and

often also in the same areas, which resulted in many duplications. Palmer also pointed out that many of the earlier workers may have naively held the notion "that most of the species were very local and that each area explored had a distinct Crataegus flora..." In this connection both Camp (1942) and Palmer (1946) pointed to the large number of species that even after several decades were known only from their type localities.

Nevertheless, Palmer (1932, 1946) and other workers (Rickett, 1936, 1937; Kruschke, 1965; Phipps & Muniyamma, 1980) have provided evidence that many, even most of the species described, particularly many of those described by Sargent, did in fact rest on real distinguishing features. The difficulty seems to have been that the earlier workers were unable to appreciate other, and perhaps more likely interpretations of their observations. As pointed out by Rickett in his studies of Missouri C. pruinosa s. l. (1936) and C. crus-galli s. l. (1937), "...the distinguishable characters are not combined at random... [as in a hybrid swarm] but are associated in a smaller number of constant combinations... It is probable that hybridization was involved in the origin of the numerous races of C. crus-galli, and probably also that their constancy and geographic restriction are results of apomictic descent from their hybrid ancestors" (Rickett, 1937).

These two features of the early concepts of Crataegus taxa, their frequent endemism and the way in which they were frequently distinguished by sets of correlated characteristics, are both seen in the data obtained here for the individual topodemes of C. crus-galli s. l. In particular, these data and the generalizations they have suggested concerning the origin and structure of Crataegus topodemes provide an explanation for the kind of species described between 1892 and 1925.

Both apomixis and self-fertility may enable production of large numbers of offspring identical to their parent. The establishment of large numbers of such offspring in the kind of sites that became increasingly abundant following the introduction of European agriculture into North America (Palmer, 1946) would have resulted in these sites having the same kind of appearance as the crus-galli sites documented here. Sympatric topodemes each made up of sibling, frequently uniparental individuals could easily have contributed to a confusion of hierarchic levels, as early workers confounded variation among groups consisting only of related individuals with variation among species.

The fact that these groups were made up of manifestly long-lived woody perennials probably also contributed to the impression that the new species to which they were referred represented real and permanent entities. Little

consideration appears to have been given to the impermanence of individual hawthorn stands (Valek, 1980) and to the possibility that they represented little more than multiple copies of a very small number of genetic individuals.

It is not true, however, that in Crataegus there are no species at all, as Van Valen (1976) has speculated. In terms of the ecological species concept proposed by Van Valen in the same article Crataegus is represented by a large number of species. These are lineages or groups of related lineages (presently described as sections, species complexes or species). Although many of these appear to share the same adaptive zone (Section 1.2, 10.6) at least temporally they have only minimally overlapping ranges, (Section 10.6.1; Appendix 7). As a result the ancestor-descendant trajectories of these species are largely independent through not only time and geographic space but also morphological (developmental) space.

Recognition of taxa within such ecologically defined "species" may be desirable as a means of indicating the presence of significant morphological discontinuities for which ecological correlates are not known. Maintenance of essentially morphological concepts of taxa for purposes of classification will nevertheless place far greater emphasis on comparisons among topodemes. This will be necessary particularly where it appears that, because of apomixis or

self-fertility, variation within topodemes is predominantly that among sibling and probably uniparental individuals.

In studies of such groups it will also become increasingly important to document morphological variation over the entire geographic range of the broadly defined groups under study. This is so because in these groups it seems unlikely that local studies can encompass enough different lineages to be able to statistically detect trends and discontinuities within the overall group. Study of taxa like C. crus-galli s. l. or the entire section Crus-galli will necessarily have to be on a continental scale and will require analytical methods that will be able to incorporate data both from existing reference collections and from newly collected topodeme samples.

11.5 Commentary on the Taxonomic Treatment of Ontario Crataegus crus-galli L. sensu lato

On the basis of the experience with Crataegus crus-galli s. l. in Ontario acquired in the course of the present study it is possible now to comment on some of the past taxonomic treatments of the group, and to make suggestions for its future treatment, specifically as it occurs in this province and by implication with regard to the complex throughout its range. Reference is made to the results of the numerical analyses described earlier

(Chapters 5-8), the data available on chromosome number (Chapter 9), and to the voucher collections that have been made, as well as to the ideas presented in this chapter and to existing treatments of the group in the literature.

11.5.1 Identification of the 10-stamen Morphotype of *Crataegus crus-galli* s. l.

As described in Section 5.3 all of the phenograms obtained for the entire 160 OTU sample on the basis of eleven reproductive descriptors distinguished all or most of the 10-stamen *C. crus-galli* OTUs from the 20-stamen ones. It is evident from the results in Chapters 6 and 8 that the morphological distinction between 10- and 20-stamen *C. crus-galli* goes far beyond stamen number (see also Appendix 7 concerning their phenological separation), and involves vegetative as well as reproductive characteristics. There is also limited documentation of a cytological contrast between the two morphotypes, in that the 10-stamen *C. crus-galli* OTUs examined were predominantly tetraploid (Table 9.2) and pollen-fertile (Fig. 10.1). The 20-stamen *C. crus-galli* OTUs studied were found to be mostly pollen-infertile triploids (Fig. 10.1; Table 9.2). These relationships are illustrated in Figure 11.1, an ordination (CVA) of the entire sample based on six flower and fruit descriptors not including stamen number.

These results confirm the way in which Phipps and Muniyamma (1980) distinguished 10-stamen C. crus-galli from the 20-stamen morphotype as C. crus-galli s. str. Their interpretation of this taxon resembles that of Kruschke (1965), as they place in synonymy with it taxa distinguished by Sargent (1908) in his account of Crataegus in Ontario (C. crus-galli var. pyracanthifolia Ait.; C. arduennae Sarg.). Thus C. crus-galli s. str. in Ontario is a somewhat variable species, including individuals with red (Sites 4, 7, 8) as well as ivory-colored (Sites 1-3, 8-10, 12) anthers, that vary in leaf characteristics as indicated in Figures 6.9, 8.1, 8.8b and d, and 8.9a and b. Phenotypic variation in C. crus-galli s. str. is also accompanied by a limited degree of uncorrelated nucleotypic variation, as a few triploid individuals have also been found (Table 9.2; Muniyamma & Phipps, 1979b).

11.5.2 Identification of the 20-stamen Morphotype of Crataegus crus-galli s. l.

Apart from differentiation of the two topodeme samples (T1, T4; Section 5.3) and a small group of OTUs with very large fruit (Section 5.4.1) the sample of 20-stamen C. crus-galli studied here is morphologically quite homogeneous. As described in Chapter 9 and Section 10.2 it appears that this sample consists principally of

pollen-infertile triploid OTUs, plus a small number of pollen-fertile tetraploid ones (compare Appendix 4 and Table 9.1). The OTUs shown to be tetraploids in the present study (OTUs 714, 751, 752 and 753) make up most of the group of 20-stamen OTUs with especially large fruit.

Examination of Figure 11.1, a CVA of the entire sample based on five *a priori* groups (the four morphotypes of *C. crus-galli* s. l. plus *C. punctata*) shows that in the space in which these groups are maximally distinct, differentiation of triploid and tetraploid 20-stamen OTUs is minimal, although it does occur along the fourth canonical variate (not shown). This axis, while it accounts for only 2.9% of the total variation in the sample, is associated with a significantly non-zero eigenvalue of $\underset{\sim}{W}^{-1}\underset{\sim}{A}$, and is most highly correlated with fruit width (WFR; Table 8.12).

Phipps and Muniyamma (1980) split the 20-stamen morphotype of *C. crus-galli* into *C. fontanesiana* (Spach) Steud. and *C. livoniana* Sarg. on the basis of the larger fruit and leaves of the latter species. Among the exemplar specimens of *C. fontanesiana* cited by them are OTUs 131 (*Phipps* 4719, 4830; Site 1) and 751 (*Barrows* 17; Site 11). Comparison of these and the other 20-stamen OTUs examined here with the exemplar specimen of *C. livoniana* cited by Phipps and Muniyamma (1980; *Phipps* 4824, 4925; Site 11) and with photographs of the type specimens and original

description of C. livoniana (Sargent, 1908) suggests that distinguishing this species from presumptive C. fontanesiana is difficult at best. Unfortunately only incomplete collections were available for the C. livoniana exemplar, so that it could not be included in the analyses described in the earlier chapters.

It may be that C. livoniana could be used to segregate the occasional tetraploid, pollen-fertile and large-fruited 20-stamen C. crus-galli individuals that are encountered in Ontario from the apparently more frequent triploid and pollen-infertile ones recognized as C. fontanesiana. However, the large fruit-sizes attained by some triploid 20-stamen OTUs, and the consequent difficulty in unequivocally distinguishing ploidy levels (especially where fresh material may be unavailable) suggests that it may be better instead to recognize only C. fontanesiana, as a somewhat variable and cytologically differentiated species, similar to C. crus-galli s. str. in these respects.

11.5.3 Putative Hybrids

Consideration elsewhere (Section 10.6; Section 11.3.5; Appendix 7) of the circumstances both contributing to and mitigating against the occurrence of hybridization between Crataegus species suggested that this process probably does occur, but that in general it probably results only

infrequently in the establishment of hybrid individuals. Proliferation of such individuals by agamospermy, self-fertility or both may nevertheless occasionally result in conspicuous numbers of individuals with a particular intermediate phenotype. This may be part of the explanation of the remaining Ontario morphotypes of C. crus-galli s. l. discussed below.

11.5.3.1 Crataegus sp. aff. C. bushii Sarg.

In Ontario this morphotype is known only from Queenston (Site 6), where it is the most frequent hawthorn species over a large and irregular area of abandoned agricultural land and the adjacent Niagara Parkway. Crataegus bushii was described by Sargent (1902) from material collected in Arkansas. In referring the Queenston material to this species, Phipps and Muniyamma (1980) suggested the possibility of its polytopic origin.

Of the four crus-galli morphotypes studied here this one departs most markedly from the usual Gestalt of C. crus-galli, by reason of its foliage (Fig. 8.9c) and the margination of its calyx lobes (Fig. 3.1d). Both of these features suggest comparison with Crategus section Macracanthae, on the following grounds. In both taxa leaves are frequently broad and rhomboidal, doubly serrate above, with conspicuously craspedodromous secondary venation (sensu

Hickey, 1973). In contrast, the leaves of C. crus-galli s. str. and C. fontanesiana tend to have a more rounded outline, and to be more lanceolate or obovate (Fig. 8.8, 8.9). In addition, the secondary venation is much more obscure and approaches being semi-craspedodromous (Hickey, 1973) as a result of frequent branching and anastomosis of the secondary veins. Finally the leaves of C. crus-galli s. str. and C. fontanesiana are much more variable with respect to the occurrence of compound teeth.

At the same time, Ontario C. sp. aff. C. bushii strongly resembles C. crus-galli s. str. and C. fontanesiana in the glossiness of the adaxial surface of its leaves and in the absence of cavities on the adaxial (radial) surface of its pyrenes (a feature characteristic of section Macracanthae). Accordingly a plausible origin for the Queenston morphotype would be a cross between the two sections. Such a hybrid origin has been proposed by Eggleston (1908a, b, 1923; Palmer, 1946, 1963) for C. prunifolia (Lam.) Pers. of European gardens, in part, as well as for a similar group of four species from the Niagara frontier and western New York State. The tetraploid chromosome number determined for one individual of C. sp. aff. C. bushii (OTU 801; Table 9.1) could have arisen in a number of ways, including fertilization of reduced gametes from tetraploid individuals of C. crus-galli s. str. (Table

9.2) and C. macracantha (Muniyamma & Phipps, 1979b).

11.5.3.2 Crataegus ?grandis Ashe

This morphotype too is known only from the Niagara peninsula (Site 4). Its appearance is intermediate between that of section Crus-galli and that of Section Punctatae because of the pubescence of its spring foliage and inflorescence axes, the glossiness of its mature leaves, and their conspicuously craspedodromous secondary venation (Hickey, 1973). As in C. punctata leaves are compoundly serrate above, with each of the secondary veins leading to a large first-order tooth bearing usually one or more smaller, higher order teeth so that leaves often appear slightly lobed above. Thorns in Ontario C. ?grandis tend to be less abundant, shorter, and more straight than in section Crus-galli, and in these features approach those of section Punctatae. Flowers of this morphotype have approximately twenty stamens with extremely small, red anthers containing at maturity no pollen at all (Plate 4c). The fruit include some of the largest observed in this study, but are more variable in size than in tetraploid C. fontanesiana. The two individuals examined cytologically here are both triploids (OTUs 601, 602; Table 9.1).

Comparable male-sterility was found by Standish (1916) in a number of taxa, several of which were subsequently also

determined to be polyploid (Longley, 1924). Such male-sterility was interpreted by Standish as an indication of hybridity. In view of the less extreme pollen infertility of triploid 20-stamen C. crus-galli (Section 10.2, Appendix 4) it is probable that in C. ?grandis some additional kind of genomic or nucleocytoplasmic incompatibility of the parents is operating (Jain, 1959).

Crataegus grandis Ashe was placed in section Punctatae by Palmer (1946, 1963) as a synonym of 10-stamen C. peoriensis Sarg. Kruschke (1965) transferred the species to section Crus-galli on account of its glossy foliage and less pubescent twigs, distinguishing it from C. peoriensis at the same time because of its stamen number. Phipps and Muniyamma (1980) included this morphotype "...somewhat arbitrarily..." within section Crus-galli, suggesting that it could represent a possible hybrid between Sections Crus-galli and Punctatae.

11.5.3.3 Other Taxa

Hybridization between Sections Crus-galli and Punctatae has been advanced as the possible origin of another taxon reported from Ontario by Phipps and Muniyamma (1980). This is C. disperma Ashe, with glabrous leaves and stems, and approximately ten stamens per flower (Ashe, 1900). Like Ashe's C. grandis, this species was placed in section

Punctatae by Palmer (1946, 1963) despite its glossy foliage; Palmer suggested that it too might represent a crus-galli x punctata hybrid (Palmer, 1963). Kruschke (1965) and Phipps and Muniyamma (1980) have treated this species under section Crus-galli, however. Phipps and Muniyamma noted the differences between it and C. ?grandis at Site 4 and suggested that it could be "...of different parental percentage and (or) later selection."

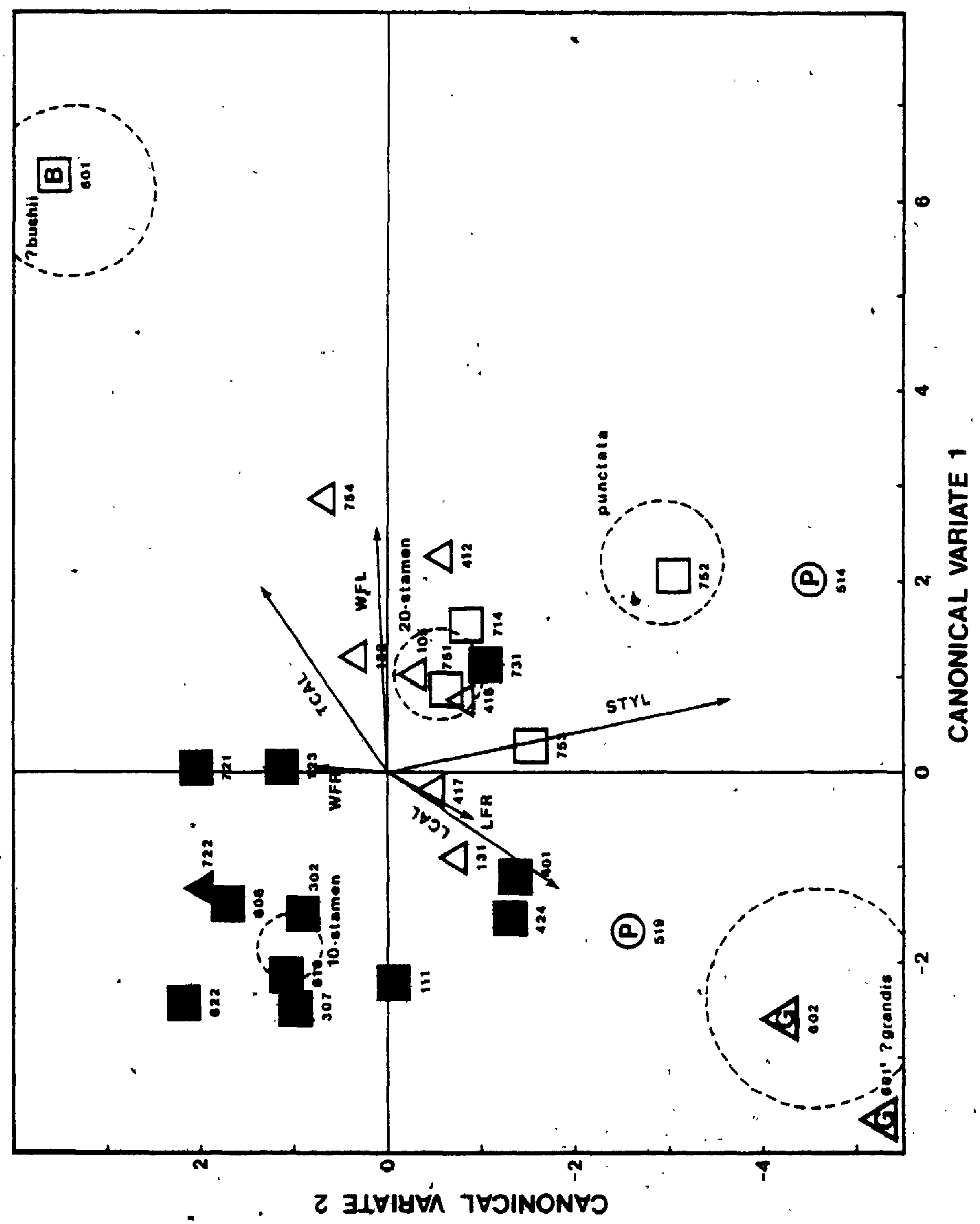
It should be noted that both of these putative hybrids between section Crus-galli and section Punctatae appear to be extremely sporadic in their occurrence in Ontario. Crategus disperma could not be included in the present study because if it is now present in the province at all, it was only re-discovered in 1982 by W. J. Crins, outside Windsor (Crins 4138). The only modern exemplar specimen of C. disperma cited by Phipps and Muniyamma (1980), OTU 761 (Stewart 2506), has since been determined to be C. macracantha, by virtue of its adaxially excavated pyrenes.

Finally, as discussed briefly already in Section 11.3.4, an hypothesis illustrated by the configuration of OTUs in Figure 11.1 is that the common 20-stamen morphotype, identified as C. fontanesiana above, represents a pseudogamously apomictic (Section 10.5) hybrid between C. crus-galli s. str. and C. punctata (Section 10.6.1), one

that is either sufficiently pollen-fertile to be able to set seed (Fig. 10.1; Fig. A3.2, Appendix 3), or succeeds in doing so at least at sites where it co-occurs with other pollen-fertile Crataegus taxa, including both 10-stamen C. crus-galli and C. punctata (Fig. 10.4). The tetraploid 20-stamen OTUs found occasionally (Table 9.1) may thus be the result of backcrosses to one or the other parent. The location of OTU 752 in Figure 11.1, compared to that of OTUs 714, 751 and 753, could be the result of backcrossing with C. punctata rather than C. crus-galli s. str.

Implicit in such a scenario is an assumption that the 10-stamen morphotype of C. crus-galli s. l. more nearly represents the ancestral form of the complex than does the 20-stamen one. An alternative hypothesis is that both common morphotypes are derived from a common ancestor without the intervention of hybridization. The significance of the former hypothesis is also limited by being based in part on ordinations in which C. punctata is the only potential parent species considered, apart from C. crus-galli s. str. However, given the frequent co-occurrence of the putative parental species (Fig. 10.4), the occasional overlap of their flowering periods (Appendix 7), and the the occurrence of other taxa of the same putative origin, the hypothesis is not implausible either.

Figure 11.1 Canonical variate analysis of 155 OTUs in five taxa belonging to Crataegus Series Crus-galli and Series Punctatae, and showing ploidy levels reported for a subsample of OTUs in Chapter 9. Canonical variates are scaled such that the mean within-group variance weighted by sample size on each variate is unity. Vectors represent contributions of the descriptors used to the canonical variates shown, in terms of the canonical weights (see Table 3.1 for explanation of descriptor abbreviations). Dashed circles show location of taxon sample centroids and approximate 99% confidence intervals around these points. Diploids are indicated by circles, triploids by triangles and tetraploids by squares. Taxa are indicated as follows: solid symbols, 10-stamen C. crus-galli s. l.; open symbols, 20-stamen C. crus-galli s. l.; B, C. sp. aff. C. bushii; G, C. ?grandis; and P, C. punctata. See Tables 8.12 and 9.1 for details.


CANONICAL VARIATE 1
CANONICAL VARIATE 2
?bushii
punctata
?grandis
20-stamen
10-stamen
STYL
TCAL
WFL
WFR
LCAL
LFR
754
412
752
514
714
731
753
417
131
424
519
602
721
722
606
302
622
307
111

APPENDIX 1

List of OTU Vouchers

Table A1.1 List of vouchers of the sample used in the present study and deposited in the Herbarium of the Department of Plant Sciences, University of Western Ontario (UWO). OTUs for which no voucher is listed were studied from sterile specimens which were not retained; they are represented by liquid-preserved material and leaf spectra in the possession of the author.

Site 1

OTU	Collector and number
101	Dickinson 709
104	Dickinson 649
105	Dickinson 696
106	Dickinson 713
108	Dickinson 796
110	Dickinson 770
111	Dickinson 652
112	Dickinson 751
114	Dickinson 585
115	Dickinson 742
116	Dickinson 575
117	Dickinson 661
118	Dickinson 717
119	Dickinson 638
120	Dickinson 790
121	Dickinson 634
122	Dickinson 600
123	Dickinson 670
124	Dickinson 760
125	Dickinson 658
126	Dickinson 794
127	Dickinson 607
129	Dickinson 726
130	Dickinson 773
131	Dickinson 662 = Phipps 4719
132	Dickinson 564
133	Dickinson 604

Table A1.1 Cont.

Site 2

OTU	Collector and number
204	Dickinson 881
207	Dickinson 884
208	Dickinson 869
209	Dickinson 875
223	Dickinson 868
224	Dickinson 860
226	Dickinson 880
401	Hopkins 093
402	Hopkins 059
403	
404	Hopkins 085
406	
407	Hopkins 078
408	Hopkins 124
409	Dickinson 547
410	Hopkins 089
411	Hopkins 076
412	Dickinson 548
414	Dickinson 549
416	Dickinson 550
417	Hopkins 061
418	Hopkins 063
419	
420	
421	Hopkins 083
422	Hopkins 102
423	Hopkins 096
424	Hopkins 122
425	Hopkins 086
426	Hopkins 074
427	Hopkins 099
428	Hopkins 071
430	Hopkins 107
752	Dickinson 950
753	Dickinson 877
754	Barrows 39
771	Dickinson 957
772	Dickinson 963

Site 3

301	Dickinson 935
302	Dickinson 920
303	Dickinson 924

Table A1.1 Cont.

Site 3 Cont.

OTU	Collector and number
304	Dickinson 927
305	Dickinson 959
306	Dickinson 931
307	Dickinson 928
308	Dickinson 929
309	Dickinson 923
310	Dickinson 916
311	Dickinson 926
312	Dickinson 917
313	Dickinson 932
314	Dickinson 918
315	Dickinson 930
316	Dickinson 934
781	Dickinson 973

Site 4

601	Dickinson 960
602	Hopkins 049
603	Dickinson 553
604	Dickinson 554
605	Dickinson 555
606	Dickinson 557
607	Dickinson 558
608	Dickinson 992
610	Dickinson 994
611	Dickinson 995
612	Dickinson 997
613	Dickinson 998
614	Dickinson 1001
615	Dickinson 1002
616	Dickinson 1003
617	Dickinson 1004
618	Dickinson 1005
619	Dickinson 1006
620	Hopkins 035
621	Hopkins 037
622	Hopkins 057
691	Dickinson 989 = Barrows 34 = Phipps 5000
692	Hopkins 045
693	Hopkins 047
694	Hopkins 051

Table A1.1 Cont.

Site 4 Cont.

OTU	Collector and number
695	Hopkins 055
696	Hopkins 053
697	Dickinson 1021

Site 5

501	
502	
503	Dickinson 898
504	
506	Dickinson 902
507	
508	
509	Dickinson 904
510	Dickinson 905
511	
512	
513	
514	Dickinson 909
516	
517	Dickinson 912
518	Dickinson 913
519	
520	
769	Dickinson 1019

Site 6

801	Dickinson 961
802	Dickinson 962
803	Hopkins 015
804	Hopkins 017
805	Hopkins 021
806	Hopkins 023
807	Hopkins 025 = Barrows 83
808	Hopkins 027
809	Barrows 82 = Phipps 5007
810	Dickinson 541
811	Dickinson 1009
812	Dickinson 1010
813	Dickinson 1015
814	Dickinson 1026
815	Dickinson 1027

Table A1.1 Cont.

OTU	Collector and number
	Site 7
791	Dickinson 1039
792	Dickinson 1042
793	Dickinson 1043
	Site 8
702	Dickinson 538
703	Dickinson 539
704	Dickinson 540
	Site 9
711	Dickinson 1029
712	Dickinson 1030
713	
714	
	Site 10
731	Dickinson 1037 = Phipps 4598
	Site 11
751	Dickinson 937 = Barrows 17
	Site 12
721	Dickinson 942 = Phipps 4604, 4637
722	Dickinson 941 = Phipps 4607a
	Site 13
761	Stewart 2506
762	Dickinson 1028

APPENDIX 2

Raw Data, Flower Fruit and Leaf Descriptors

Table A2.1 Raw Data, Flower and Fruit Descriptors. (See Table 3.1 for explanation of descriptor abbreviations and units of measurement; see Table 5.1 for the site affiliation of each OTU).

Table A2.1 Cont.

OTU	STYL	STAM	PROJ	TCAL	PUB1	WFL	LCAL	LFR	WFR	PUB2	ANTH
104	2.20	18.05	1.00	0.70	0.00	4.45	4.28	11.55	10.80	0.00	0.00
105	2.13	17.72	1.00	0.67	0.00	4.55	4.40	11.03	11.25	0.00	1.00
116	2.15	18.70	0.89	0.80	0.00	3.83	3.94	11.23	10.68	0.00	0.00
117	2.00	17.16	0.89	0.89	0.00	4.30	4.05	12.18	12.50	0.00	0.00
119	2.10	18.50	1.00	0.80	0.00	4.20	4.40	11.55	10.85	1.00	0.00
122	2.30	18.70	1.00	0.95	0.00	4.10	4.40	11.45	11.40	1.00	0.00
127	2.05	18.50	0.91	0.95	0.00	4.40	4.50	11.30	11.10	0.00	0.00
131	2.20	16.65	0.70	0.55	0.00	3.90	4.25	11.38	11.85	0.00	0.00
132	2.15	16.95	1.00	1.10	0.00	4.15	4.40	10.29	10.55	0.00	0.00
133	1.75	17.15	1.00	0.67	0.00	4.55	4.25	10.70	10.30	0.00	1.00
101	1.32	8.37	0.00	1.00	0.35	3.55	5.00	10.05	11.18	1.00	0.00
106	1.35	8.45	0.00	0.90	0.20	3.35	4.70	10.80	12.13	1.00	0.00
108	1.40	8.50	0.14	1.10	0.15	3.33	4.56	10.08	10.83	1.00	0.00
110	1.95	7.60	0.20	0.30	0.00	3.20	3.90	12.15	12.28	1.00	0.00
111	1.83	9.41	0.00	0.80	0.00	3.40	5.15	10.58	11.37	1.00	0.00
112	1.70	8.10	0.00	1.16	0.32	3.46	5.00	10.13	10.80	2.00	0.00
114	1.75	9.55	0.00	1.16	0.11	3.20	4.67	10.35	11.13	1.00	0.00
115	1.95	8.60	0.00	0.85	0.70	3.25	5.05	10.40	11.63	1.00	0.00
118	1.40	8.15	0.17	1.35	0.00	3.71	4.29	12.45	12.10	0.00	0.00
120	1.80	8.85	0.00	0.80	0.95	3.25	5.00	10.32	11.05	1.00	0.00
121	1.65	8.90	0.00	1.15	0.20	3.44	5.17	9.95	11.10	1.00	0.00
123	1.77	8.18	0.00	1.15	0.00	3.84	4.83	10.18	10.90	1.25	0.00
124	1.55	9.00	0.00	1.00	0.00	3.15	4.90	9.83	9.00	1.00	0.00
125	1.50	8.70	0.00	0.90	0.20	3.60	5.35	10.08	10.70	2.00	0.00
126	1.50	8.80	0.00	1.00	0.35	3.29	4.71	10.35	11.08	2.00	0.00
129	1.25	8.50	0.00	1.00	0.05	3.40	4.65	10.28	11.23	2.00	0.00
130	1.20	8.05	0.00	0.63	0.20	3.36	5.27	10.17	10.17	2.00	0.00
301	1.30	7.10	0.00	0.58	0.00	3.55	4.50	11.70	11.15	0.00	0.00
302	1.35	7.55	0.00	0.75	0.00	3.90	5.05	11.59	10.67	0.00	0.00
303	1.53	7.50	0.00	1.10	0.00	4.10	5.20	10.88	9.88	0.00	0.00
304	1.25	7.10	0.00	0.80	0.00	3.65	4.45	11.20	9.78	0.00	0.00
305	1.26	6.68	0.00	1.11	0.00	3.44	4.72	10.50	9.48	0.00	0.00

Table A2.1 Cont.

OTU	STYL	STAM	PROJ	TCAL	PUB1	WFL	LCAL	LFR	WFR	PUB2	ANTH
306	1.11	6.89	0.00	0.84	0.00	3.61	4.06	9.92	9.50	0.00	0.00
307	1.40	7.05	0.00	0.75	0.00	3.39	4.94	10.45	9.93	0.00	0.00
308	1.05	5.95	0.00	0.75	0.00	3.15	4.85	10.53	9.93	0.00	0.00
309	1.15	8.55	0.00	1.15	0.00	3.60	4.00	11.00	10.33	0.00	0.00
310	1.21	6.32	0.00	1.11	0.00	3.40	5.00	10.18	9.53	0.00	0.00
311	1.55	7.50	0.00	0.60	0.00	3.44	4.44	9.75	9.53	0.00	0.00
312	1.45	6.45	0.00	0.65	0.00	3.55	4.65	9.90	9.55	0.00	0.00
313	1.35	7.15	0.00	0.95	0.00	3.70	5.25	10.60	10.70	0.00	0.00
314	1.70	7.55	0.00	0.65	0.00	3.50	4.72	10.20	10.15	0.00	0.00
315	1.20	7.25	0.00	0.80	0.00	3.20	4.50	9.83	8.83	0.00	0.00
316	1.30	7.35	0.00	0.80	0.00	3.40	4.45	11.00	10.65	0.00	0.00
751	2.05	16.30	1.00	0.47	0.00	4.95	4.80	13.13	14.90	2.00	2.00
752	2.44	18.90	0.90	0.51	0.03	5.78	6.40	13.72	17.14	1.00	2.00
753	2.13	17.47	1.00	0.44	0.00	4.96	5.42	13.65	16.18	0.00	1.00
754	2.15	19.20	1.00	1.30	0.00	4.80	4.85	11.70	14.08	0.00	2.00
224	2.10	18.05	1.00	0.90	0.00	4.40	4.40	12.63	14.28	0.00	1.00
204	2.27	19.55	0.50	0.45	0.00	5.05	5.15	13.35	15.33	1.00	0.92
207	2.10	17.80	1.00	0.72	0.00	4.55	4.20	12.28	14.48	0.00	1.00
208	2.05	18.35	1.00	0.95	0.00	4.35	4.45	12.00	14.04	0.00	1.00
209	2.25	18.00	1.00	0.80	0.00	4.10	4.10	11.50	13.17	1.00	1.00
402	2.15	18.30	1.00	0.65	0.00	4.65	5.00	12.28	13.93	0.00	1.00
403	2.20	18.50	1.00	0.70	0.00	4.65	4.60	12.08	13.29	0.00	1.00
404	2.40	19.25	1.00	0.90	0.00	4.85	5.30	11.83	13.93	0.00	1.00
406	2.60	18.90	1.00	0.79	0.05	4.61	4.78	11.88	13.85	0.00	1.00
407	2.30	18.40	1.00	0.65	0.05	4.80	5.35	11.55	13.17	0.00	1.00
408	2.30	18.20	1.00	0.71	0.05	4.85	4.83	12.18	14.05	0.00	1.00
409	2.40	18.90	1.00	1.10	0.00	4.80	4.72	11.83	13.83	0.00	1.00
410	2.10	17.15	1.00	0.70	0.00	4.70	4.55	11.83	13.75	0.00	1.00
411	2.20	18.80	1.00	0.75	0.00	4.60	4.95	11.90	13.65	0.00	1.00
412	2.40	18.30	1.00	0.84	0.00	4.85	4.45	11.59	13.42	0.00	1.00
414	2.20	18.65	1.00	0.70	0.00	4.55	4.41	11.73	13.55	0.00	1.00
417	2.20	17.95	1.00	0.65	0.00	4.20	4.50	12.33	14.33	0.00	1.00

Table A2.1 Cont.

OTU	STYL	STAM	PROJ	TCAL	PUB1	WFL	LCAL	LFR	WFR	PUB2	ANTH
418	2.30	19.55	1.00	0.75	0.00	4.45	4.50	12.30	13.33	0.00	1.00
419	2.37	17.26	1.00	0.50	0.00	4.93	5.29	10.75	13.75	0.08	1.00
420	2.25	18.10	1.00	0.75	0.00	4.60	4.70	12.03	14.03	0.00	1.00
422	2.15	18.90	1.00	0.70	0.00	4.80	4.70	11.23	13.10	0.00	1.00
423	2.10	18.15	1.00	0.75	0.00	4.55	4.75	10.87	12.76	0.00	1.00
425	2.25	19.15	1.00	0.90	0.00	4.80	5.10	11.88	13.40	0.00	0.00
426	2.25	18.50	0.90	0.75	0.00	4.70	4.70	11.78	14.05	0.00	0.00
427	2.21	18.15	1.00	0.90	0.00	4.55	5.10	12.63	13.83	0.00	1.00
223	2.05	10.90	0.10	0.30	0.00	3.95	4.45	13.53	13.30	1.00	0.00
226	2.00	9.75	0.00	0.44	0.00	3.86	4.05	13.63	13.00	0.00	0.00
401	2.20	11.40	0.00	0.47	0.00	4.10	4.60	12.93	13.23	1.00	0.00
416	1.95	11.75	0.00	0.30	0.15	3.95	4.45	13.23	13.25	1.00	0.00
421	2.05	11.15	0.00	0.47	0.00	4.00	4.35	11.98	13.88	0.00	0.00
424	2.05	10.45	0.00	0.45	0.00	4.10	5.05	12.65	13.35	1.00	0.00
428	1.75	9.85	0.00	0.55	0.00	3.95	4.00	12.68	13.42	1.00	0.00
430	1.65	10.20	0.00	0.40	0.00	4.05	4.70	13.03	12.68	1.00	0.00
771	3.03	13.90	0.00	0.87	3.00	3.69	5.05	10.53	12.15	3.00	2.00
772	3.15	18.10	0.00	0.95	3.00	5.07	4.30	16.03	17.95	3.00	2.00
781	3.37	19.10	0.17	0.78	3.00	5.31	5.46	12.30	15.98	3.00	0.00
501	2.55	19.30	0.00	0.35	3.00	4.40	4.35	11.57	12.97	3.00	2.00
502	2.49	17.69	0.50	1.70	3.00	4.15	6.25	11.19	11.58	3.00	2.00
503	3.21	19.51	0.00	1.55	3.00	4.35	4.95	11.55	12.55	3.00	3.00
504	3.47	18.75	0.29	1.05	2.95	4.70	4.20	12.25	13.38	3.00	2.00
506	3.35	19.65	0.88	0.82	3.00	4.50	5.60	11.03	11.81	3.00	1.00
507	3.20	19.25	0.00	1.42	3.00	4.55	5.20	12.08	13.78	3.00	2.00
508	2.85	19.60	0.25	0.95	3.00	4.55	5.70	13.83	12.93	3.00	2.00
509	2.85	17.55	0.80	1.15	3.00	4.25	4.80	12.20	12.75	3.00	3.00
510	3.58	18.37	1.00	1.44	3.00	4.65	5.35	12.10	14.63	3.00	3.00
511	3.20	19.10	0.40	1.65	3.00	4.40	6.15	11.40	12.20	3.00	2.00
512	3.15	19.00	0.70	1.75	3.00	4.45	4.40	11.98	14.88	3.00	2.00
513	2.90	19.25	0.25	1.90	3.00	4.35	4.50	12.18	13.57	3.00	2.00
514	3.50	19.10	0.00	1.30	2.60	4.35	5.75	10.85	11.93	3.00	2.00

Table A2.1 Cont.

OTU	STYL	STAM	PROJ	TCAL	PUB1	WFL	LCAL	LFR	WFR	PUB2	ANTH
516	3.05	19.85	1.00	1.85	3.00	4.75	6.45	13.71	14.72	3.00	3.00
517	3.05	18.90	0.13	1.30	3.00	4.38	4.71	11.83	14.28	3.00	3.00
518	3.95	20.05	0.10	2.35	3.00	4.70	5.50	12.75	14.00	3.00	3.00
519	2.42	19.11	0.71	0.00	3.00	4.28	4.33	13.85	14.50	3.00	2.00
520	3.17	19.37	0.25	0.68	3.00	4.70	5.45	11.35	13.80	3.00	2.00
601	2.65	19.35	0.44	0.10	2.70	3.95	6.45	13.94	14.87	2.00	3.00
602	2.64	19.15	0.47	0.41	2.51	3.98	5.95	14.85	15.78	3.00	3.00
691	2.51	19.18	0.59	0.45	2.44	4.05	6.24	13.10	15.20	3.00	3.00
692	2.38	19.75	0.75	0.56	2.94	4.44	6.50	13.49	14.62	2.00	3.00
693	2.75	19.10	0.58	0.50	2.95	4.31	7.00	13.63	15.14	3.00	3.00
694	2.80	19.80	0.20	0.25	3.00	4.15	6.15	14.22	15.00	2.00	3.00
695	2.65	19.60	0.60	0.55	3.00	4.10	6.80	14.16	14.91	3.00	3.00
696	1.70	12.10	0.09	0.35	0.00	4.25	4.70	13.61	12.77	1.00	2.00
697	1.95	9.75	0.00	0.80	0.00	3.60	4.10	13.55	13.38	0.00	2.00
603	1.05	9.45	0.00	0.35	0.00	3.78	4.83	10.65	9.35	0.00	3.00
604	0.90	9.25	0.00	0.20	0.00	3.75	4.30	10.93	9.90	0.00	3.00
605	0.90	9.05	0.00	0.45	0.00	3.68	4.36	11.38	10.98	0.00	3.00
606	1.20	9.55	0.00	0.75	0.00	3.95	4.75	12.38	11.80	0.12	3.00
607	1.00	9.60	0.00	0.40	0.00	3.85	5.00	11.75	11.25	0.00	3.00
608	0.95	8.90	0.00	0.15	0.00	3.80	4.45	12.13	11.05	1.00	3.00
610	1.15	9.65	0.00	0.53	0.00	4.00	5.15	12.35	11.98	0.00	3.00
611	1.00	9.55	0.00	0.40	0.00	3.90	4.90	12.65	11.43	0.00	3.00
612	1.15	9.75	0.00	0.40	0.00	3.90	4.55	12.50	11.78	0.00	3.00
613	1.30	9.60	0.00	0.80	0.00	4.00	5.25	12.65	12.33	0.00	3.00
614	1.00	9.05	0.00	0.20	0.00	3.85	4.40	12.53	11.38	0.00	3.00
615	0.85	9.20	0.00	0.35	0.00	3.95	4.50	11.21	10.68	0.00	3.00
616	1.10	9.40	0.00	0.30	0.00	4.05	4.60	10.80	9.95	0.00	3.00
617	0.95	9.35	0.00	0.25	0.00	3.90	4.45	12.38	11.30	0.00	3.00
618	1.05	8.75	0.00	0.40	0.00	3.75	4.50	12.85	12.23	0.00	3.00
619	1.10	9.60	0.00	0.65	0.00	3.85	5.00	11.83	11.13	1.00	3.00
620	1.20	9.60	0.00	0.50	0.00	3.80	4.65	11.30	10.93	0.00	3.00
621	1.00	9.40	0.00	0.25	0.00	3.85	4.35	11.17	11.06	0.00	3.00

Table A2.1 Cont.

OTU	STYL	STAM	PROJ	TCAL	PUB1	WFL	LCAL	LFR	WFR	PUB2	ANTH
622	0.95	9.25	0.00	0.30	0.00	3.90	4.30	11.93	11.60	0.00	3.00
802	2.03	10.86	0.11	0.71	0.00	3.89	4.11	12.10	12.55	2.00	0.00
807	2.05	10.95	0.20	0.53	0.10	3.80	4.00	10.81	11.38	0.00	0.00
808	2.13	10.67	0.10	0.67	0.00	3.65	4.05	11.85	12.10	1.00	0.00
814	2.00	9.40	0.00	0.21	0.00	3.60	3.95	11.35	11.78	2.00	0.00
801	1.96	13.39	0.00	3.00	0.00	4.56	5.25	11.73	12.93	1.00	2.00
803	1.90	13.20	0.00	3.00	0.00	4.65	4.90	11.44	11.39	1.00	1.00
804	1.95	11.95	0.10	3.00	0.00	4.45	5.10	11.28	12.15	0.00	1.00
805	1.90	10.35	0.00	3.00	0.00	4.24	5.00	11.00	11.70	1.00	2.00
806	2.11	12.63	0.00	3.00	0.00	4.77	5.27	10.54	12.31	0.00	2.00
809	1.85	12.60	0.00	3.00	0.00	3.78	5.78	10.51	12.42	1.00	2.00
810	2.10	10.15	0.00	3.00	0.00	4.20	4.95	10.99	11.71	2.00	2.00
811	1.97	11.79	0.00	3.00	0.00	4.68	5.45	11.73	13.93	1.00	1.00
812	2.00	10.05	0.00	3.00	0.05	4.60	5.70	11.38	12.75	0.90	2.00
813	1.95	11.28	0.00	2.97	0.00	4.85	6.15	11.30	12.23	0.00	1.00
815	2.30	9.80	0.00	3.00	0.00	4.50	5.40	11.63	12.55	2.00	2.00
702	1.85	9.80	0.56	0.25	0.11	3.60	4.30	11.10	11.30	1.00	3.00
703	2.05	9.70	0.20	0.17	0.00	3.55	4.45	11.18	12.00	2.00	0.00
704	1.15	8.30	0.00	0.10	0.00	3.55	4.25	11.65	11.13	0.00	3.00
711	2.25	10.30	0.00	0.06	0.05	3.55	3.25	11.35	11.88	1.00	0.00
712	2.05	12.40	0.00	0.63	0.00	4.15	4.22	12.10	12.98	1.00	0.00
713	2.10	16.25	0.78	0.25	0.00	4.85	4.50	12.29	14.89	1.00	2.00
714	2.05	18.65	0.92	0.35	0.00	5.35	5.00	12.43	15.13	1.00	1.00
721	1.45	7.80	0.00	1.00	0.00	3.95	4.30	10.63	10.10	0.00	0.00
722	1.20	7.65	0.00	0.60	0.00	4.00	4.45	11.15	11.05	0.00	0.00
731	2.35	11.85	0.00	1.10	0.55	4.35	4.95	12.33	12.33	2.00	0.00
761	2.75	5.90	0.11	3.00	2.00	4.23	6.73	8.00	9.85	3.00	1.00
762	3.15	6.30	0.10	3.00	1.80	4.45	7.80	8.03	10.05	3.00	1.00
769	2.50	19.65	0.00	3.00	2.75	3.67	4.93	8.94	10.39	3.00	2.00
791	1.75	10.00	0.63	0.75	0.00	3.75	4.79	10.41	9.88	0.00	2.00
792	1.90	9.15	0.90	1.40	0.00	3.45	4.80	10.43	10.63	1.00	2.00
793	1.55	9.90	0.83	1.00	0.00	3.63	4.81	10.93	10.70	0.00	2.00

Table A2.2 Raw data, short shoot terminal leaf descriptors. See Table 3.3 for explanation of descriptor abbreviations and units of measurement. See Table 5.1 for site affiliation of each OTU.

OTU	X	Y	Z	NUMSEC	ANGSEC	TEETH
104	16.25	17.63	44.13	6.38	31.25	6.38
116	19.00	16.14	35.86	5.57	25.00	6.00
117	23.17	21.50	46.50	6.00	30.00	5.83
119	21.50	20.13	47.38	5.63	28.13	5.25
122	19.38	18.38	41.38	6.00	25.63	6.75
127	20.75	17.50	40.63	5.63	30.00	6.38
131	20.43	17.29	42.86	6.14	29.29	6.43
133	20.14	20.86	48.29	6.43	27.86	5.86
101	17.25	16.88	39.38	7.13	32.50	7.13
108	17.88	19.38	41.88	6.63	30.63	7.63
115	16.00	15.75	34.25	7.13	33.75	7.75
120	16.88	16.88	42.50	7.50	32.50	6.63
121	16.88	18.00	42.38	6.75	30.00	6.38
123	16.71	19.14	37.86	7.00	34.29	6.86
124	16.77	18.77	38.23	6.85	37.31	7.69
129	19.00	20.00	44.38	7.63	35.00	6.50
301	15.71	19.57	35.00	9.00	40.00	7.14
305	16.00	22.63	38.00	9.38	36.25	7.88
306	17.14	20.29	37.43	8.71	40.00	8.00
307	16.20	19.07	37.00	9.07	40.00	7.93
310	13.71	19.71	37.86	8.43	35.71	8.57
311	12.14	14.14	29.14	6.86	40.00	8.00
312	12.71	17.57	33.43	7.14	39.29	8.14
313	17.29	20.14	35.00	7.86	37.14	7.14
204	20.67	25.00	45.83	7.00	30.00	8.17
208	15.17	19.50	37.00	5.00	30.00	7.67
209	20.83	22.92	43.08	6.17	30.42	7.33
403	17.57	21.00	43.71	6.43	31.43	7.71
404	17.17	21.17	38.67	6.50	31.67	7.67
407	16.75	20.63	38.63	6.25	28.75	7.38
410	18.29	22.14	39.57	5.86	27.86	7.00
417	20.57	22.43	44.14	6.71	32.14	6.43
418	16.00	20.00	37.00	6.14	30.00	8.43
423	20.71	23.57	49.71	6.14	28.57	6.00

Table A2.2 Cont.

OTU	X	Y	Z	NUMSEC	ANGSEC	TEETH
226	22.86	21.43	37.00	7.00	40.00	5.57
401	21.86	21.43	40.86	7.29	31.43	5.86
416	21.71	20.29	38.57	7.57	40.00	5.71
421	18.43	18.29	34.57	6.57	38.57	6.14
424	22.75	22.13	37.63	7.13	40.00	6.00
428	22.67	20.83	35.83	8.17	38.33	6.00
430	22.20	21.80	36.40	7.00	36.00	5.40
606	20.38	22.25	40.63	8.25	38.13	6.75
608	21.88	21.75	39.63	8.50	40.63	7.00
610	22.50	23.38	43.00	8.75	38.75	6.75
612	17.14	19.86	36.86	7.43	37.14	6.71
616	19.38	22.75	39.38	8.38	39.38	5.88
617	18.63	20.75	39.75	7.88	37.50	6.88
620	17.43	20.29	35.86	7.00	41.43	7.00
621	16.63	18.38	35.63	7.25	36.25	6.50
622	22.43	22.57	41.43	7.86	43.57	6.43
503	18.29	23.00	40.00	6.57	30.00	9.71
506	20.67	24.83	46.50	6.83	30.00	8.83
508	19.80	22.40	45.20	6.20	30.00	8.80
511	19.67	22.17	42.33	7.00	30.00	9.00
512	16.60	19.60	41.80	7.60	30.00	11.20
514	17.67	19.00	40.50	7.33	28.33	10.00
517	21.57	28.29	56.14	7.43	31.43	10.86
520	19.75	23.88	49.75	7.63	25.63	11.13
601	18.00	23.13	44.38	6.50	29.38	7.63
602	19.13	22.13	45.50	6.38	29.38	7.75
801	19.69	30.92	49.69	6.54	33.85	9.85
806	23.50	35.00	53.17	6.83	31.67	9.17
811	26.40	31.20	46.60	7.60	37.00	8.40
813	19.13	25.63	38.25	6.25	31.25	11.38
815	18.75	26.38	40.50	6.50	37.50	10.13

APPENDIX 3

Reproductive Morphology

A3.1 Introduction

Careful, and where possible, quantitative description of reproductive structures and their development are necessary in order to be able to draw significant conclusions about patterns of reproductive phenomena. The descriptions of reproductive structures that follow here, as well as the descriptions of their development in Sections 10.2 and 10.3 are primarily qualitative, however, since available resources were concentrated on other aspects of the study.

A3.2 Inflorescence

Judging from the structures present in overwintering buds at the end of the winter (Rickett, 1943; Plate 7), and from reports on related genera (Church, 1908; Luckwill, 1970), inflorescences are initiated in the summer of the year preceding their opening. Inflorescences are terminal on short shoots which develop from and whose growth is continued by an axillary bud. The inflorescence itself was characterized by Rickett as a cyme; following his redefinition of this inflorescence type as "a more or less

flat-topped compound cluster composed essentially of dichasia or pleiochasia" (Rickett, 1943; compare Fig. 6 and 7 of Phipps and Muniyamma, 1980).

Flower number per inflorescence varies among the taxa dealt with here. Based on observations made in the course of the pollination experiments described below (Section 10.5), the modal range varies from 10-14 flowers in inflorescences of C. punctata to 20-24 flowers, in C. sp. aff. C. bushii. The common 10- and 20-stamen morphotypes both have a modal range of 15-19 flowers per inflorescence, but the distributions (based on 210 and 193 inflorescences, for five and four individuals respectively) are skewed in opposite directions. The 20-stamen morphotype has the lower mean number of flowers per inflorescence. Mean flower number in the two individuals of C. ?grandis studied experimentally is even closer to that of C. punctata.

It was uncommon, however, for all the flowers produced on an inflorescence to reach anthesis, let alone set fruit. It appears common that approximately a third or more of the flowers are shed before opening, in the crus-galli morphotypes.

In the present study it was not possible to estimate the number of inflorescences produced by individual trees. It seems likely, though, that individuals of a size

frequently encountered, say 3-4 m high and with a crown almost as broad produce hundreds of inflorescences in a good year.

A3.3 Floral Structure

The flowers of C. crus-galli s. l. and C. punctata are 1.5-2.0 cm across the fully open corolla in Ontario material. Their flowers, like those of the genus as a whole, are actinomorphic and 5-merous. Considerable variation in number of parts may occur, however, in the corolla and, within each morphotype, in the androecium and gynoecium as well (Table 3.11).

Perianth and androecium are borne on the rim of a hypanthial cup (Fig. A3.1; compare Decaisne, 1874; Koehne, 1890). Development of this cup apparently involves growth beneath primordia not only of the perianth and androecium, but also the abaxial portions of the individual gynoecial primordia (Plate 7; Kania, 1973), as in Cotoneaster (Payer, 1857), Cydonia and Pyrus (Church, 1908). As a result the abaxial walls of the locules are formed by the hypanthium while their radial walls and the styles develop from the gynoecial primordia (Plate 7d, e; Church, 1908). The mature flower thus consists of a hypanthial (inferior) ovary (Leins, Merxmüller & Sattler, 1972) the upper, exposed surface of which forms a nectariferous disk surrounding the syles (D, Fig. A3.1).

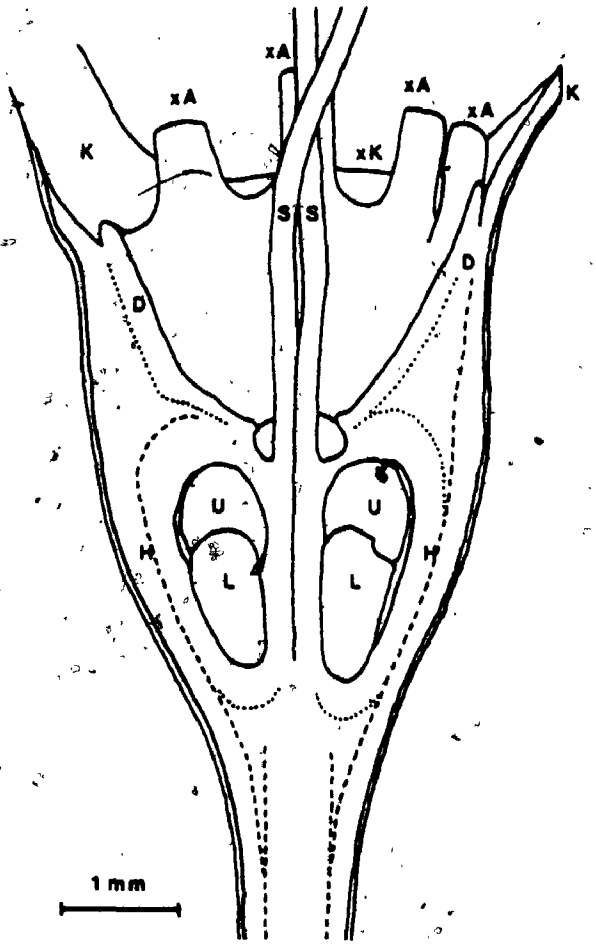


Figure A3.1 Flower of 10-stamen Crataegus crus-galli longitudinally bisected to show gynoecial construction. The corolla has been completely removed. Explanation of labelling: K, calyx lobes (xK, lobe removed); xA, stamen removed; D, hypanthial disk; H, outer wall of hypanthial ovary; S, style; and upper (U) and lower (L) ovules. See Plates 4 and 8 for details of ovule morphology and placentation.

Plate 7 Early floral development in 10-stamen Crataegus crus-galli (a-c and d-e at the scales indicated): (a) corolla initiation, between calyx lobes; (b) early petal (lower left) and stamen (upper right) primordia; (c) petal primordia alternating with calyx lobes; and (d, e) longitudinally bisected flowers showing developing gynoecial primordium with style suture (arrows) and hypanthial disk. In (c) the lowermost petal, and in (d) the right side of the flower were damaged during dissection. Explanation of labelling: Br, subtending bract; K, calyx lobes; C, petals; A, stamens; G, free portion of the gynoecial primordium which will develop into style and stigma; and D (hatched), hypanthial disk. Collection data: (a,b) OTU 731, 10.4 London ° days, 19 April 1979; (c) OTU 755, 46.6 London ° days, 7 May 1978; and (d, e) OTU 731. 59.6 London ° days, 2 May 1979.


a
Br
b
0.1 mr
c
K
d
C
C
0.1
e

A3.3.1 Perianth

Calyx lobes are usually elongate-triangular (Fig. 3.1a-e), varying in length (Table 3.1). The use of calyx lobe margination as a descriptor (TCAL) in the numerical studies reflects the great variation in degree of toothing observed, among the morphotypes of C. crus-galli s. l. and within the sample of C. punctata.

White petals approximately 7-10 mm long alternate with the calyx lobes, and resemble those illustrated by Dau (1941) for material referred to C. oxyacantha L. Although predominantly imbricate, their aestivation varies considerably, as described by Church (1908) for Cydonia japonica Pers. (= Chaenomeles japonica (Pers.) Lindl.) In some individuals petals often enclose groups of stamens in such a way as to isolate them from the style.

A3.3.2 Androecium

Besides differentiation of the two common morphotypes of C. crus-galli s. l. according to stamen number, anther color and the presence of a projection of the anther connective ("distal appendage," Hufford, 1980) also vary among the taxa studied (Table 3.11). Anthers are dorsifixed and dehisce introrsely. Dehiscence does not usually occur until after flowers open and the filaments have reflexed

from the center of the flower. Anther dehiscence in unopened flowers, possibly leading to bud autogamy has been observed only occasionally, in the course of the pollination experiments.

A3.3.3 Gynoecium

The stigma in both the crus-galli and punctata material examined is capitate, approximately 0.5 mm in diameter. It appears to be of the wet papillate type described by Heslop-Harrison and Shivanna (1977) for some genera of the Maloideae (Malus, Mespilus, Pyrus, Raphiolepis). Papillae are approximately 30 µm in diameter (Plate 8b) and approximately 30 - 35 µm long (Plate 4e).

Styles vary in length from five to ten mm among samples. Since each style results from a single gynoecial primordium, style number is an indicator of locule number. Each locule contains almost invariably two superposed ovules; occasionally only one was found, and sometimes three. Each ovule is anatropous, bitegmic and crassinucellate (Plate 8a, c, d; Plate 5, 6; Péchoutre, 1902). The two ovules are contrasted in that the lower has a short funiculus and a long micropyle, opposed to an obturator. The upper ovule has a long funiculus and a short micropyle, and no obturator (Plate 8a, c, d; Plate 4f; Péchoutre, 1902; Sterling, 1964; Muniyamma & Phipps, 1979a).

Plate 8 Gynoecial morphology of Crataegus punctata (a; OTU 512) and 20-stamen C. crus-galli (b-d; OTU 752): superposed ovules within a single locule (a, c, d) and papillate stigmatic surface (b) of recently opened flowers (all at the same scale). Explanation of labelling: L, lower ovule; U, upper ovule; O, obturator; and m, position of the micropyle. Funiculus of the lower ovule is hidden behind that of the upper one.


a
L
b
c
m
d
U
m
0.5mm
L
m
O

A3.4 Fruit and Seed

The fruit which develops from the flower of Crataegus was characterized by early workers in terms of its similarity to an apple fruit as a pome with bony endocarps (Lindley, 1822; Decaisne, 1874). Koehne (1890) distinguished the Maloideae with bony endocarps as the tribe Crataegeae Koehne; the other genera in the subfamily with cartilaginous endocarps constituted tribe Sorbeae Koehne. The fruit of the Crataegeae have been described by Kalkman (1973) in terms of their composition as polypyrenous drupes.

In both C. crus-galli s. l. and C. punctata mature fruit are red (or yellow, in C. punctata var. aurea; Fig. 1a-d, f-h, Phipps & Muniyamma, 1980), but this is often obscured by scabbing so that accurate determinations of fruit color are frequently difficult or impossible. The fruit flesh (pericarp) of C. crus-galli s. l. is relatively dry, tart, and somewhat astringent due to the presence of abundant phenolic compounds.

Only a single seed develops within the pyrene, characteristically from the lower ovule in the material examined (M. Shaik, personal communication). Earlier workers (Decaisne, 1874; Koehne, 1890) described the upper ovule as sterile, but this seems an overstatement in view of the apparently normal embryo-sac development that takes

place in upper ovules prior to anthesis (Section 10.3). What instead may be critical is the separation of the micropyle of the upper ovule from the obturator (Plate 5a, c, d). If pollen tubes entering the locule are directed first to the obturator, then only the micropyle of the lower ovule is nearby.

Péchoutre (1902) and Corner (1976) note that seeds of the Crataegeae are characterized by thinner and unlignified testas, compared to those of the Sorbeae in which the endocarp does not lignify. The mature seed of Crataegus consists of a single linear embryo enclosed by a layer of endosperm, plus tegmen and testa (Fig. 42, 43, Péchoutre, 1902; Corner, 1976; Fig. 2, Brinkman, 1974).

Crataegus seed require imbibition at summer temperatures followed by cold stratification in order to germinate (Fordham, 1960). Where the endocarp is massive, as in the C. crus-galli s. l. and C. punctata germination in the field may be delayed until the second spring following seed-set (Flemion, 1938). More rapid germination may be obtained either by removing seed from the endocarp and stratifying it as recommended by Fordham (1960), or else by scarifying the intact pyrenes with sulfuric acid prior to stratification (Flemion, 1941; Brinkman, 1974).

Parthenocarpy, that is the development of mature fruit containing exclusively seedless pyrenes, was found to be infrequent in open-pollinated fruit except at Site 4 (Figure A3.2).

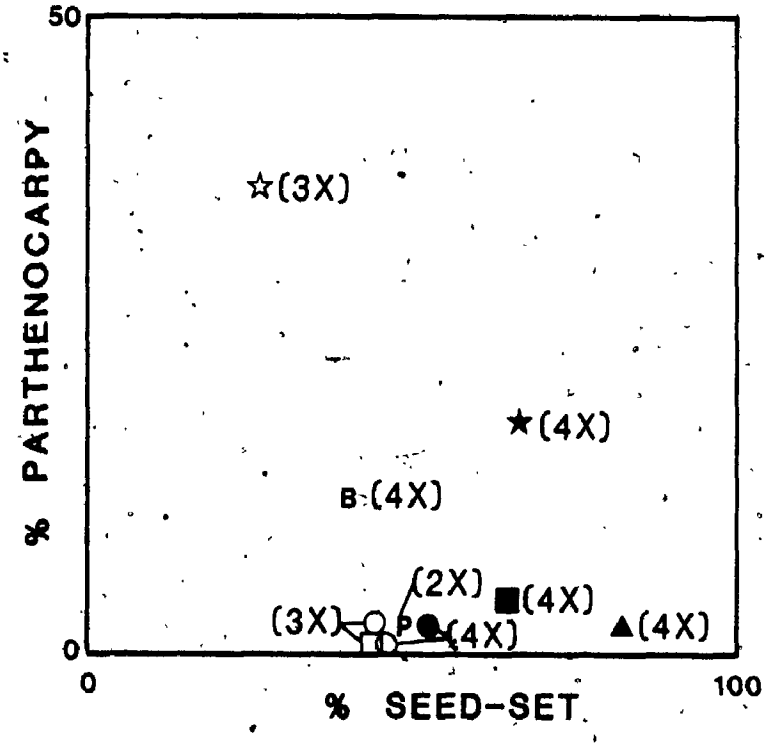


Figure A3.2 Frequency of parthenocarpy in relation to seed-set in open-pollinated fruit of the Crataegus individuals studied experimentally (Section 10.5). Symbols representing topodeme samples as described for Figure 10.1. Ploidy level is indicated for each individual or group of individuals, based on the data in Table 9.1. Frequency of parthenocarpy is calculated as percent fruit for which seed-score was zero. Frequency of seed-set calculated as mean percent seed-score per fruit.

APPENDIX 4

Pollen Data

(Compiled by M. M. Macleod, unpublished reports, 1980, 1981)

Table A4.1 Raw data, pollen grain size. Data are counts of pollen grains with equatorial diameters in in the 0.1 eyepiece unit classes indicated (1 eyepiece unit = 32µm), and mean pollen grain diameter in this plane. See Table 9.1 for details concerning ploidy level and site.

OTU	ploidy	1.0	1.1	1.2	1.3	1.4	1.5	mean diameter (µm)
10-stamen *C. crus-galli*								
722	3X	0	0	8	21	6	5	43.6
(1)	3X	0	0	0	7	11	17	48.7
111	4X	0	6	7	4	10	10	44.1
123	4X	0	0	13	13	4	5	43.1
302	4X	0	23	8	0	0	0	36.9
307	4X	0	0	6	16	19	0	43.7
721	4X	0	0	7	14	15	6	41.6
731	4X	0	0	9	11	12	9	44.7

Table A4.1 Cont.

OTU	ploidy	1.0	1.1	1.2	1.3	1.4	1.5	mean diameter (μm)
20-stamen C. crus-galli								
131	3X	0	0	6	16	13	0	43.8
132	3X	0	0	8	12	13	4	43.2
752	4X	0	0	3	5	16	13	46.6
C. sp. aff. C. bushii								
811	(2)	0	0	0	17	19	1	44.9
815	(2)	0	0	0	16	18	0	45.0
C. punctata								
514	2X	14	16	2	4	0	0	36.0
519	2X	2	22	23	0	0	0	38.1

(1) Phipps 4454, Site 1 (determined to be a triploid by Muniyamma and Phipps, 1979b).

(2) No data available.

Table A4.2 Raw data, pollen grain number and stainability. Explanation of Abbreviations: STAIN, % stainable pollen grains; PGPER, average number of pollen grains per anther; POLOVU, pollen-ovule ratio. See Section 10.2 for further explanation. See Table 5.1 for further identification of OTUs.

OTU.	STAIN	PGPER	POLOVU
10-stamen *C. crus-galli*:			
101	88.97	1494.0	4229.4
111	90.02	2091.3	4835.3
123	90.19	2690.7	4943.5
123	89.23	1888.7	3431.9
(1)	89.42	2200.3	4964.1
302	89.31	2519.0	6291.7
304	88.98	2498.7	6312.4
307	87.44	2069.3	4555.6
312	89.13	2319.3	4598.0
401	80.75	1762.3	3687.7
424	68.93	2315.0	4067.9
603	94.28	2143.7	9093.0
604	92.82	1886.7	8996.5
606	93.57	1939.7	7220.9
606	93.09	1678.7	5682.4
608	92.57	1982.7	8594.0
610	94.33	2256.7	8931.1
613	92.36	2513.3	8571.1
616	94.54	2244.0	9066.7
619	94.29	2076.0	8544.0
622	94.07	1865.7	8542.5
702	86.36	1877.3	4294.3
711	91.51	1828.3	3830.1
721	82.97	1753.7	3912.6
722	84.68	1827.3	4932.1
731	86.89	1694.0	3779.6
791	66.06	1764.3	3329.5
(2)	87.03	1094.7	2230.1
696	60.13	2020.7	4144.2
697	48.65	1654.7	2011.7

Table A4.2 Cont.

OTU	STAIN	PGPER	POLOVU
20-stamen C. crus-galli:			
105	10.90	2751.0	1389.5
122	20.44	2032.3	1688.4
131	34.56	2116.3	2767.4
132	19.30	2144.0	1721.2
133	16.22	1525.3	1211.9
208	46.57	1982.3	4132.5
211	87.68	1726.3	6174.1
403	24.59	1924.3	1990.2
406	28.30	2110.3	2171.1
408	23.76	1439.0	1353.1
412	22.42	1814.7	1550.4
418	19.45	2218.3	1833.2
426	22.67	1414.7	1318.3
714	62.20	1834.7	5188.6
(3)	37.32	1722.7	2733.6
751	92.45	2065.3	7823.6
752	85.73	1732.3	5807.7
754	24.89	1845.3	2051.0
755	21.84	1303.7	1160.0
C. punctata:			
501	97.59	2267.7	8373.4
507	34.73	1683.7	1758.6
508	46.20	2660.3	4227.2
512	94.64	2873.3	8201.2
514	95.12	2748.7	7134.3
516	95.47	3329.7	10343.7
519	82.82	3295.3	10770.7
C. conspecta:			
(4)	88.52	2532.0	4624.1
(4)	85.22	2147.0	3623.4
C. sp. aff. C. bushii:			
801	98.70	1904.3	6419.5
811	97.12	2753.0	7036.8
813	96.72	2215.7	5196.0

Table A4.2 Cont.

(1) Dickinson 871b, Site 2.

(2) Dickinson 1036, Site 10; used as source of pollen for experiments performed on OTU 731 (Section 10.5).

(3) Dickinson 1020, Ron Smith woodlot, Concession 12, London Township.

(4) Dickinson 999 and 1000, Site 4; used as source of pollen for experiments performed on OTU 602 (Section 10.5).

APPENDIX 5

Biotic Interactions

A5.1 Introduction

Crataegus reproduction is conspicuously affected by interactions with other organisms at a number of stages, from the time shoots and inflorescences expand until seedlings become established. In addition to the flower visitors described in Section 10.4 (Table A5.1), a large number of other insects either feed on or parasitize *Crataegus* reproductive structures (Section A5.2; Table A5.1; Wellhouse, 1922). *Crataegus* is also the host of the aecial stages of the rust fungus *Gymnosporangium* Hedwig f. ex DC. Severe infestation by this fungus galls young shoots so that flower buds are not produced, and vegetative growth is reduced (Hoover, 1961). Infected fruits are usually sterile. Finally, as described in Section A5.3, animals are implicated in the dispersal of *Crataegus* species.

There are a number of features of hawthorns that influence these interactions. In the case of *C. crus-galli* s. l., as in the genus as a whole, the flowers are morphologically unspecialized but emit a characteristic odor, partly sweet and partly urine-like, that appears to be attractive to small bees and flies (Knuth, 1908; Proctor &

Yeo, 1973). Nectar is produced by the perigynous disk. As described in Section 10.4 pollen production may be copious when unimpaired by triploidy. Wainio and Brooks (1941) give the following values for the composition of the fruit of Pennsylvania C. crus-galli (as percent dry weight, of the pericarps presumably): crude protein, 2.8; crude fat, 3.3; crude fiber, 32.8; total carbohydrates, 57.4; phosphorus, 0.04; and calcium, 0.42 Individual seeds appear to have a much higher lipid and protein content.

A5.2. Insects Feeding on Reproductive Structures

Of the large insect fauna associated with hawthorns (Wellhouse, 1922), many species feed on parts of the reproductive structures, causing varying degrees of interference with seed-set. One species was especially common on 10- and 20-stamen C. crus-galli at some of the sites studied here (Table A5.1; Power, unpubl.): Anthonomus nebulosus (Coleoptera: Curculionidae).

Eggs of A. nebulosus are laid within the corolla of immature flowers. They hatch after about a week (Wellhouse, 1922), and the larva then remains within the closed corolla feeding on the anthers. Pupation and emergence occur after the larva has fed for a couple of weeks (Wellhouse, 1922), during which time the flower fails to open. Pollination may apparently occur during this time since fruit of

C. crus-galli s. l. could sometimes be found with the u[illegible]ened, empty corolla intact. In general, however, infestation of inflorescences by larvae of A. nebulosus resulted in taking a considerable number of flowers out of possible seed-production.

A5.3 Dispersal

Dispersal of Crataegus propagules is as yet poorly understood. On the one hand, many birds and mammals in North America are known to eat Crataegus fruit (Martin, Zim & Nelson, 1951), and pyrenes containing viable seed apparently can be recovered from the droppings of at least some of these animals. On the other hand, observation of poorly managed agricultural land or other suitable sites (Section 1.2) indicates that dispersal certainly occurs. What is not known is what behavior, by which animals, is most frequently responsible for this dispersal. Fruit of C. crus-galli s. l. tends to remain fresh throughout the winter, either on the tree or on the ground, so that effective dispersal may take place under a variety of different conditions.

Table A5.1 Insects found in association with 10- and 20-stamen C. crus-galli at Sites 2 and 3. Based on data collected and reported by N. Power (unpubl., 1980). Numbers in parentheses are Power collection numbers; specimens are lodged in UWO.

TAXON	SITE	ROLE
COLEOPTERA		
Cantharidae		
Cantharis rectis (72, 99)	2, 3	Nectar feeder
C. bilineatus (82a)	3	" "
Scarabidae		
Trichiotinus cf. *piger* (85)	3	Nectar & pollen feeder
Curculionidae		
Anthonomus nebulosus (88, 90, 94)	2, 3	Inflorescence feeder
Polydrusus sp. (74)	2	" "
(subfam. Brachyrinidae; 70)	2, 3	Foliage feeder
LEPIDOPTERA		
Tortricidae		
Pandemis limitata (125)	3	Larvae found feeding on flower buds bound together

Table A5.1 Cont.

TAXON	SITE	ROLE
Archips argyrospilla (107, 110)	2, 3	Larvae feed on leaves and inflorescences
Olethreutes chionosema (103)	3	Larvae feed on leaves and inflorescences
Lycaenidae		
(subfam. Theclinae; 111)	3	General feeder?
Geometridae		
_____ (140)	2, 3	Inflorescence and foliage feeder
Noctuidae		
Catocala crataegi (129)	2	Foliage feeder; rare
Lasiocampidae		
Malacosoma sp. (127)	2, 3	" "
DIPTERA		
Syrphidae		
cf. Didea fasciata (1)	3	Nectar & pollen feeder
Sarcophagidae		
cf. Sarcophaga haemorrhoidalis (3)	3	Nectar feeder
Calliphoridae		
Lucila sp. (6)	3	" "

Table A5.1 Cont.

TAXON		SITE	ROLE
Muscidae			
_____	(7)	3	Nectar feeder
_____	(8)	3	" "
_____	(12)	2	" "
cf. Bombylidae			
_____	(14)	2	" "
HYMENOPTERA			
Apidae			
Bombus cf. vagans vagans	(49)	3	Nectar & pollen feeder
Apis mellifera	(53, 60)	2, 3	" " " "
_____	(61)	2	" " " "
Formicidae			
_____	(38)	3	Nectar feeder
_____	(43)	3	" "
_____	(58)	2	" "

APPENDIX 6

Pollination Experiment Data

Table A6.1 Results of pollination experiments performed on individuals of Crataegus crus-galli s. l. and C. punctata. See Section 10.5 for details of experimental treatments (A - F). See Table 10.3 for complete identification of the OTUs studied. Statistics (1 - 3) calculated from arcsin-square root transformed data.

(1) Back-transformed mean percent seed-score.

(2) Back-transformed 95% confidence interval of mean percent seed-score.

(3) Back-transformed range, percent seed-score.

(4) (Number of trees)/(Number of replicates)/(Number of treated styles).

* OTUs 412, 418, 754 (triploid individuals).

** OTU 752 (tetraploid individual).

*** Pollen from Crataegus conspecta Sarg. (Dickinson 999 and 1000).

Table A6.1 Cont.

		Treatment				
		A	B	C	D	E
20-stamen C. crus-galli:						
Site 1	(1)	0.95	0	0.03	0.16	0
	(2)	0.06-2.87		-0.04-0.30	-0.05-1.02	
	(3)	0 - 10.94		0 - 5.00	0 - 12.50	
	(4)	3/14/312	3/11/222	3/13/186	3/14/166	3/5/96
Site 2*	(1)	1.00	0.01	0	0	4.13
	(2)	0.22-2.35	-0.01-0.10			0.46-11.23
	(3)	0 - 11.43	0 - 3.45			0 - 50.00
	(4)	3/19/577	3/18/405	3/18/282	3/17/263	3/19/325
Site 2**	(1)	6.78	7.85	1.03	24.83	1.39
	(2)	1.16-16.55	3.32-14.09	0-4.14	14.05-37.50	-0.05-6.51
	(3)	0 - 22.73	2.70-16.18	0 - 10.00	11.11-39.13	0 - 7.14
	(4)	1/8/299	1/7/309	1/8/310	1/6/181	1/6/144
10-stamen C. crus-galli:						
Site 1	(1)	3.73	1.07	0	24.31	27.20
	(2)	0.21-11.30	0-4.11		13.89-36.56	16.81-39.04
	(3)	0 - 17.65	0 - 14.29		0 - 55.56	5.56-50.00
	(4)	2/9/164	2/12/189	2/12/129	2/12/143	2/12/116
Site 2	(1)	15.33	1.93	0	10.97	21.34
	(2)	5.17-29.60	0.15-5.67		2.22-25.16	12.11-32.34
	(3)	0 - 43.48	0 - 19.05		0 - 44.45	0 - 50.00
	(4)	2/9/218	2/12/217	2/9/94	2/12/126	2/12/131

Table A6.1 Cont.

		Treatment A	B	C	D	E
10-stamen *C. crus-galli* cont.:						
Site 3	(1)	1.68	2.10	0.02	0.93	5.81
	(2)	0.27-4.24	0.49-4.78	-0.02-0.17	0-3.52	1.44-12.83
	(3)	0 - 12.50	0 - 16.67	0 - 5.00	0 - 11.11	0 - 22.73
	(4)	3/18/185	3/21/178	3/17/131	2/11/98	2/12/110
Site 4	(1)	0.68	0.22	0.23	5.04	0.95
	(2)	0-2.56	0-1.36	0-1.42	0.69-13.04	0.04-3.02
	(3)	0 - 8.33	0 - 7.14	0 - 7.69	0 - 33.33	0 - 6.25
	(4)	2/12/232	2/11/194	2/11/144	2/12/147	2/11/143
Site 10	(1)	5.37	5.83	0.05	17.22	19.10
	(2)	0-38.05	0.92-14.57	0-0.66	5.14-34.42	5.34-38.69
	(3)	0-22.23	0-12.50	0-1.85	4.17-31.25	7.14-40.0
	(4)	1/4/105	1/6/165	1/6/152	1/5/100	1/5/100
C. sp. aff. *C. bushii*:						
Site 6	(1)	8.62	14.07	0	17.66	11.13
	(2)	1.90-19.52	11.71-16.58		7.35-31.22	4.10-21.03
	(3)	0 - 22.22	6.82-19.23		8.00-28.57	0 - 26.67
	(4)	2/7/151	2/12/350	2/12/172	1/5/79	2/8/131
C. punctata:						
Site 5	(1)	14.63	0.04	0.01	0.02	24.51
	(2)	9.76-20.30	-0.01-0.22	-0.01-0.07	-0.03-0.24	11.27-40.87
	(3)	0 - 34.00	0 - 4.17	0 - 2.44	0 - 3.33	0 - 66.67
	(4)	3/23/717	3/18/672	3/19/442	2/12/235	2/11/246
C. ?*grandis*:					Treatment F***	
Site 4	(1)	0.01	0	0	22.09	
	(2)	-0.02-0.14			4.78-42.27	
	(3)	0 - 1.96			0 - 40.00	
	(4)	2/12/359	2/12/359	2/12/441	1/11/248	

APPENDIX 7

Phenology of Flowering

A7.1 Introduction

A conspicuous feature of Crataegus taxa is the pattern of flowering phenology that they exhibit. As noted by Sargent (1907) and other early workers (e.g. Lecoq & Lamotte, 1847) sympatric entities which are morphologically distinct also tend to flower at different times. Differentiation of the 10- and 20-stamen morphotypes of C. crus-galli s. l. in the field is often possible, where they occur together, on the basis of phenology alone during the spring. Such differences in the time of anthesis among sympatric entities represents a potential mechanism for reproductive isolation. For reasons such as these, systematic documentation of phenological differences among Crategus entities is desirable. Phipps and Muniyamma (Table 2, 1980) and Smith, Phipps and Dickinson (1980) have provided semi-quantitative descriptions of the characteristic flowering periods of Ontario Crataegus taxa, based on herbarium records. The present section provides more detailed documentation of the phenological differences among some of the taxa studied here, based on repeated field observations at a limited number of sites during April, May

and June, 1978-1981.

A7.2 Method

In climatic regions with pronounced seasonal variation in temperature, plant growth is usually also seasonal, occurring only when temperatures exceed some minimum value. The timing of growth events which are initiated or resume when such minimum values are exceeded is often better characterized (for greater generality and predictivity) by an appropriate climatic parameter than by calendar date. Such a parameter is the temperature sum (Sarvas, 1967; Reader, 1975). This is the number of degrees by which daily mean temperatures (DMT) exceed a threshold value required for growth activity, accumulated from the beginning of the growing season (accumulated degree-days).

Temperature sums were employed by Phipps and Muniyamma (1980) and Smith et al (1980), using a threshold value of 5 C. This value is used here also, since casual observations suggest that negligible development (bud opening, shoot and inflorescence expansion) takes place when DMT remain below this temperature. Temperature sum accumulations over the entire growing season have been mapped for Ontario (Webber & Hoffman, 1967); these maps have been used in selecting sites (Chapter 2) and interpreting distributions (Chapter 11).

The temperature sums used here are based on DMT recorded by the Atmospheric Environment Service, Environment Canada, at London, Ontario (London Airport). Phenological events are related to the temperature sum accumulated up to the calendar date one day prior to their observation, on the principle that predictive use of these data (e. g. to facilitate long-range collecting; Phipps & Muniyamma, 1980) requires correlation of the events on day x with the heat accumulated through day (x - 1).

Flowering period was estimated subjectively. However, since most estimates were related to when the pollination-experiments described in Section 10.5 would be possible, it seems likely that these estimates provide a satisfactory picture of prevailing conditions. For a given date sites or individual trees were estimated as being at early-, mid- or late anthesis, according to the relative proportions of open and unopened flowers remaining. Mid-anthesis was considered as the stage at which half of the flowers in most inflorescences appeared to have opened.

A7.3 Results

Figure A7.1 summarizes flowering period observations at Sites 1-5 during 1978-1981. In general, 20-stamen C. crus-galli tends to flower (reach mid-anthesis) 20 or more degree-days (2-7 days) before sympatric 10-stamen

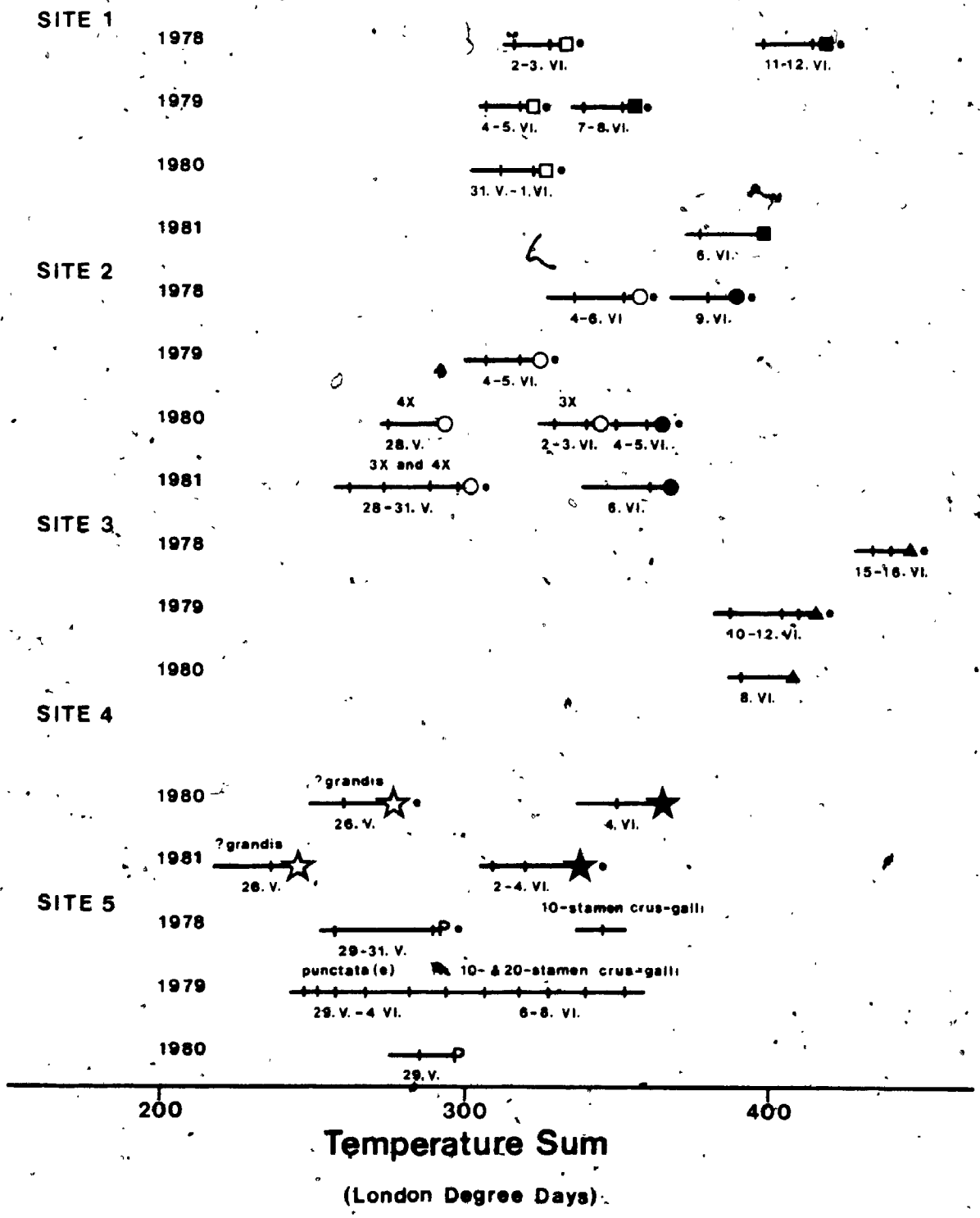
C. crus-galli. The flowering periods of C. ?grandis and 10-stamen C. crus-galli at Site 4 differ by approximately 60 degree-days (7-9 days). The separation between the flowering periods of C. punctata and the 10- and 20-stamen morphotypes of C. crus-galli is usually of the same order of magnitude both at Site 5 (Fig. A7.1) and at the other sites where the two species occur together (not shown).

Further differentiation with respect to flowering period was noticed among both 20-stamen C. crus-galli and C. punctata at Site 2. In the former case this corresponded to the morphological differentiation (Chapter 5) of tetraploid individuals (OTUs 751-753) from triploid ones (Chapter 9). These tetraploid individuals tended to flower slightly earlier than the triploid ones. In the latter case, an individual of C. punctata (OTU 771) at Site 2 consistently flowered much later than other punctata individuals there, 1979 through 1981.

It appears that when the advance of spring is retarded by periods of cold weather subsequent accumulation of heat takes place rapidly, so that increases in temperature sum which take, say, five days or more in most years take place in only two or three, following a cold spell. Such a telescoping effect can result in considerable overlap of otherwise well-separated flowering periods. In 1979 only 49.4 degree-days accumulated between May 21 and 31, as

compared with 84.1 during the preceding ten days, and 120.1 during the following ten. At Site 5, individuals of C. punctata were pollinated from May 29 to June 3. On June 6 both 10- and 20-stamen C. crus-galli individuals were found adjacent a punctata one on which flowers with undehisced anthers still remained. Thus pollen transfers may be possible from time to time between individuals of species or morphotypes whose flowering periods do not usually overlap at a given site.

Figure A7.1 Flowering phenology of Crataegus taxa at Sites 1 through 5, 1978-1981. Horizontal lines indicate estimated flowering period. Cross-ticks indicate the temperature sum accumulated up to the dates on which observations were made. Open symbols, 20-stamen C. crus-galli topodeme samples (T1, T4); solid symbols, 10-stamen C. crus-galli topodeme samples (T2, T3, T5, T6); individual symbols as defined in Table 3.2. Data for the topodeme sample of C. punctata at Site 5 (T7) indicated by the letter P. Additional individuals of these and other taxa as indicated. Records based on the experiments described in Section 10.5 indicated by the letter e.


SITE 1
1978
2-3. VI.
11-12. VI.
1979
4-5. VI.
7-8. VI.
1980
31. V.-1. VI.
1981
6. VI.
SITE 2
1978
4-6. VI
9. VI.
1979
4-5. VI.
1980
4X
28. V.
3X
2-3. VI.
4-5. VI.
1981
3X and 4X
28-31. V.
6. VI.
SITE 3
1978
15-16. VI.
1979
10-12. VI.
1980
8. VI.
SITE 4
1980
?grandis
26. V.
4. VI.
1981
?grandis
26. V.
2-4. VI.
SITE 5
10-stamen crus-galli
1978
29-31. V.
punctata (e)
10- & 20-stamen crus-galli
1979
29. V.-4 VI.
6-8. VI.
1980
29. V.
200
300
400
Temperature Sum
(London Degree Days)

REFERENCES

Alexander, M. P. 1969. Differential staining of aborted and non-aborted pollen. Stain Technol. 44:117-122.

Anderson, T. W. 1958. An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, Inc., New York.

Anderson, T. W. 1963. Asymptotic theory for principal component analysis. Ann. Math. Stat. 34:122-148.

Andrews, D. F. 1972. Plots of high-dimensional data. Biometrics 28:125-136.

Ashby, E. 1950. Studies in the morphogenesis of leaves. VI. Some effects of length of day upon leaf shape in Ipomoea caerulea. New Phytol. 49:375-387.

Ashe, W. W. 1900. Crataegus disperma. J. Elisha Mitchell Sci. Soc. 17:14-15.

Asker, S. 1979. Progress in apomixis research. Hereditas 91:231-240.

Atchley, W. R., Gaskins, C. T. & D. Anderson. 1976. Statistical properties of ratios. I. Empirical results. Syst. Zool. 25:137-148.

Babcock, E. B. & G. L. Stebbins, Jr. 1938. The American species of Crepis. Carnegie Inst. Washington Publ. No. 504.

Bailey, L. H. 1949. Manual of Cultivated Plants. Macmillan Co., New York.

Baker, H. 1953. Race formation and reproductive method in flowering plants. Symp. Soc. Exp. Biol. 7:114-145.

Bărа, I. I. & G. I. Gniorgiță. 1973. Individual variability of certain characters in sexual and apomictic species of the genus Ranunculus. L. S. "Stejarul," Ecol. Ter. Gen. 1972-1973:103-114. [in Rumanian with English abstract]

Bărа, I. I., Gniorgiță, G. I. & E. G. Totn. 1973. Comparative study of the biology of populations of the apomictic species Ranunculus auricomus L. L. S. "Stejarul," Ecol. Ter. Gen. 1972-1973:83-101. [in

Rumanian with English abstract].

Barcikowski, R. S. & J. P. Stevens. 1975. A Monte Carlo study of the stability of canonical correlations, canonical weights and canonical variate - variable correlations. Multivariate Behavioral Res. 10:353-364.

Barrett, S. C. H. 1982. Genetic variation in weeds. Pages 73-98 in R. Charudattan & H. Walker [eds.], Biological Control of Weeds with Plant Pathogens. John Wiley & Sons, Inc., New York.

Bartlett, M. S. 1937. Properties of sufficiency and statistical tests. Proc. Roy. Soc. London Ser. A. 160:268-282.

Bartlett, M. S. 1947. Multivariate analysis. J. Roy. Stat. Soc. Ser. B. 9:176-197.

Bayer, R. J. & G. L. Stebbins, Jr. 1981. Chromosome numbers of North American species of Antennaria Gaertner (Asteraceae: Inuleae). Amer. J. Bot. 68:1342-1349.

Beadle, C. D. 1903. Crataegus. Pages 532-569 in J. K. Small [ed.], Flora of the Southeastern United States. By the editor, New York.

Bennett, M. D. 1971. The duration of meiosis. Proc. Roy. Soc. London Ser. B. 178:277-299.

Berlyn, G. P. & J. P. Miksche. 1976. Botanical Microtechnique and Cytochemistry. Iowa State University Press, Ames, Iowa.

Beshir, M. E. 1975. A multivariate analysis of character variation in Jack Pine. Ph.D. Thesis. University of Western Ontario.

Beyer, W. H. [ed.] 1968. Handbook of Tables for Probability and Statistics. Chemical Rubber Co., Cleveland, Ohio.

Blackburn, D. T. 1980. A generalized distance metric for the analysis of variable taxa. Bot. Gaz. (Crawfordsville) 141:325-335.

Blackith, R. E. & R. A. Reyment. 1971. Multivariate Morphometrics. Academic Press, London & New York.

Box, G. E. P. 1949. A general distribution theory for a class of likelihood criteria. Biometrika 36:317-346.

Bradfield, G. E. & L. Orloci. 1975. Classification of vegetation data from an open beach environment in southwestern Ontario: cluster analysis followed by generalized distance assignment. Canad. J. Bot. 53:495-502.

Bradshaw, A. D. 1953. Human influences on hybridization in Crataegus. Pages 181-183 in J. E. Lousley [ed.] The Changing Flora of Britain. T. Buncle, for the Bot. Soc. Br. Is., Arbroath.

Bradshaw, A. D. 1971. The significance of hawthorns. Pages 20-29 in Standing Conference for Local History [ed.], Hedges and Local History. National Council of Social Service, London.

Bradshaw, A. D. 1975. Crataegus L., Crataegus L. x Mespilus L. = Crataemespilus Camus. Pages 230-232 in C. A. Stace [ed.] Hybridization and the Flora of the British Isles. Academic Press, London.

Bremermann, H. J. 1980. Sex and polymorphism as strategies in host-pathogen interactions. J. Theor. Biol. 87:671-702.

Bremermann, H. J. undated preprint. Towards a theory of sex. I. A new model. Center for Pure and Applied Mathematics, University of California, Berkeley. PAM-19.

Brinkmann, K. A. 1974. Crataegus L. Pages 356-360 in C. S. Schopmeyer [Technical Coordinator, U. S. F. S.], Seeds of Woody Plants in the United States, Agriculture Handbook No. 450. Forest Service, U. S. Department of Agriculture, Washington, D. C.

Brown, H. B. 1910. The genus Crataegus with some theories of the origin of species. Bull. Torrey Bot. Club 37:251-260.

Brown, M. B. & A. B. Forsythe. 1974. The small sample behavior of some statistics which test the equality of several means. Technometrics 16:129-132.

Byatt, J. I. 1975. Hybridization between Crataegus monogyna Jacq. and C. laevigata (Poiret) DC. in southeastern England. Watsonia 10:253-264.

Byatt, J. I. 1976a. The structure of some Crataegus populations in northeastern France and southeastern Belgium. Watsonia 11:105-115.

Byatt, J. I. 1976b. The genus Crataegus in Greece. Candollea 31:283-301.

Byatt, J. I. 1976c. Pollen morphology of some European species of Crataegus L. and of Mespilus L. (Rosaceae). Pollen et Spores 18:335-349.

Byatt, J. I., Ferguson, I. K. & B. G. Murray. 1977. Intergeneric hybrids between Crataegus L. and Mespilus L.: a fresh look at an old problem. J. Linn. Soc., Bot. 74:329-343.

Byatt, J. I. & B. G. Murray. 1977. Chromosome numbers of some Eurasian species of Crataegus. Watsonia 11:374-375.

Calder, J. A., Taylor R. L. & G. A. Mulligan. 1968. Flora of the Queen Charlotte Islands. Pt. 2, Cytological Aspects of the Vascular Plants. Canada Department of Agriculture Research Branch Monograph 4(2).

Camp, W. H. 1942. The Crataegus problem. Castanea 7:51-55.

Camp, W. H. & C. L. Gilly. 1943. The structure and origin of species. Brittonia 4:323-385.

Campbell, N. A. 1976. A multivariate approach to variation in microfilariae: Examination of the species Wuchereria lewisii and demes of the species W. bancrofti. Austral. J. Zool. 24:105-114.

Campbell, N. A. 1978. Multivariate analysis in biological anthropology: some further considerations. J. Human Evol. 7:197-203.

Campbell, N. A. 1979. Some practical aspects of canonical variates analysis. BIAS 6:7-18.

Campbell, N. A. 1981. Graphical comparisons of covariance matrices. Austral. J. Stat. 23:21-37.

Campbell, N. A. & R. J. Mahon. 1974. A multivariate study of variation in two species of Rock Crab of the genus Leptograspus. Austral. J. Zool. 22:417-425.

Campbell, N. A. & W. R. Atchley. 1981. The geometry of canonical variate analysis. Syst. Zool. 30:268-280.

Catling, P. M. 1982. Breeding systems of northeastern North American Spiranthes (Orchidaceae). Canad. J. Bot. 60:3017-3039.

Challice, J. & M. Kovanda. 1978. Chemotaxonomic survey of the genus Sorbus. Naturwissenschaften 65:111.

Chan, L. S. 1972. Treatment of missing values in discriminant analysis, I. Sampling experiment. J. Amer. Stat. Assoc. 67:473-477.

Charlesworth, B. 1980. The cost of sex in relation to mating system. J. Theor. Biol. 84:655-671.

Cheng, P. C., Walden, D. B. & R. I. Greyson. 1979. Improved plant microtechniques for TEM, SEM, and LM specimen preparation. Nat. Sci. Council Monthly, ROC 7:1000-1007.

Church, A. H. 1908. Types of Floral Mechanism, Part I, Types I-XII. Clarendon Press, Oxford.

Clark, G. [ed.] 1973. Staining Procedures used by the Biological Staining Commission, 3rd ed. Williams & Wilkins Co., Baltimore, Maryland.

Clausen, J. 1954. Partial apomixis as an equilibrium system in evolution. Caryologia 6 (Suppl.):469-479.

Cooley, W. W. & P. R. Lonnes. 1971. Multivariate Data Analysis. John Wiley & Sons, Inc., New York.

Coombs, E. A. K. in press. Apomixis in the Poa secunda complex. Amer. J. Bot.

Cormack, R. M. 1971. A review of classification. J. Roy. Stat. Soc. A. 134:321-367.

Corner, E. J. H. 1976. The Seeds of the Dicotyledons, Vol. 1. Cambridge University Press, Cambridge.

Crovello, T. J. 1968. The effect of alteration of technique at two stages in a numerical taxonomic study. Univ. Kansas Sci. Bull. 47:761-786.

Cruden, R. W. 1977. Pollen-ovule ratios: a conservative indicator of breeding systems in flowering plants. Evolution 31:32-46.

Cruden, R. W. & S. Miller-Ward. 1981. Pollen-ovule ratio, pollen size, and the ratio of stigmatic area to the pollen-bearing area of the pollinator: an hypothesis. Evolution 35:964-974.

Darlington, C. D. 1939. The Evolution of Genetic Systems. Cambridge University Press, Cambridge.

Darlington, C. D. & L. F. LaCour. 1976. The Handling of Chromosomes, 6th Ed. Allen & Unwin, Ltd., London.

Dau, M. 1941. Der Weissdorn Crataegus oxyacantha L. Ph.D. Thesis. Hanseatic University, Hamburg.

Davis, G. L. 1966. Systematic Embryology of the Angiosperms. John Wiley & Sons, Inc., New York.

Davis, M. B. 1976. Pleistocene biogeography of temperate deciduous forests. Geosciences and Man 13:13-26.

Davis, P. H. & V. H. Heywood. 1963. Principles of Angiosperm Taxonomy. Oliver & Boyd, Edinburgh & London. [reprinted with corrections and supplementary bibliography, 1965]

Day, B. & R. A. Fisher. 1937. The comparison of variability in populations having unequal means. Ann. Eugen. 7:333-348.

Decaisne, M. J. 1874. Mémoire sur la famille des Pomacées. Nouv. Arch. Mus. Hist. Nat. 10:113-192, plates 8-15.

Dixon, W. J. & M. B. Brown [eds.]. 1979. BMDP-79 Biomedical Computer Programs P-Series. University of California Press, Berkeley.

Dupree, A. H. 1959. Asa Gray 1810-1888. Harvard University Press, Cambridge, Mass.

Eggleston, W. W. 1908a. The Crataegi of the northeastern United States and adjacent Canada. Rhodora 10:73-84.

Eggleston, W. W. 1908b. Crataegus L. Pages 460-479 in B. L. Robinson & M. L. Fernald [eds.], Gray's New Manual of Botany, 7th Ed. American Book Co., New York.

Eggleston, W. W. 1910. Sketch of the Crataegus problem, with special reference to work in the south. J. N. Y. Bot. Gard. 11:78-83.

Eggleston, W. W. 1923. Crataegus L. Page 63 in Report of the State Botanist for 1921. N. Y. State Mus. Bull. 243-244.

Eldred, R. A. & M. A. Maun. 1982. A multivariate approach to the problem of decline in vigor of Ammophila. Canad. J. Bot. 60:1371-1380.

El-Gazzar, A. 1980. The taxonomic significance of leaf

morphology in Crataegus (Rosaceae). Bot. Jahrb. Syst. 101:457-469.

Faegri, K. & J. Iversen. 1964. Textbook of Pollen Analysis. Munksgaard, Copenhagen.

Faegri, K. & L. van der Pijl. 1979. Principles of Pollination Ecology, 3rd Ed. Pergamon Press, Oxford.

Fagerlind, F. 1940. Die Terminologie der Apomixis-Prozesse. Hereditas 26:1-22.

Farris, J. S. 1969. A successive approach to character weighting. Syst. Zool. 18:374-385.

Feder, N. & T. P. O'Brien. 1968. Plant microtechnique: some principles and new methods. Amer. J. Bot. 55:123-142.

Feoli, E. & L. Orlóci. 1979. Analysis of concentration and detection of underlying factors in structured tables. Vegetatio 40:49-54.

Fernald, M. L. 1950. Gray's Manual of Botany, 8th Ed. American Book Co., New York.

Fewster, P. & L. Orlóci. 1978. Stereograms to aid group recognition and trend identification in vegetation data. Canad. J. Bot. 56:162-165.

Flake, R. H., von Rudloff, E. & B. L. Turner. 1969. Quantitative study of clinal variation in Juniperus virginiana using terpenoid data. Proc. Natl. Acad. U. S. A. 64:487-494.

Flemion, F. 1938. A rapid method for determining the viability of dormant seeds. Contr. Boyce Thompson Inst. 9:339-351.

Flemion, F. 1941. Further studies on the rapid determination of the germinative capacity of seeds. Contr. Boyce Thompson Inst. 11:455-464.

Fordham, A. J. 1960. Propagation of woody plants by seed. Arnoldia 20:33-40.

Gilmartin, A. J. 1974. Variation within populations and classification (I). Taxon 23:523-536.

Gilmartin, A. J. 1980. Variation within populations and classification. II. Patterns of variation within

Asclepiadaceae and Umbelliferae. Taxon 29:199-210.

Gilmour, J. S. L. & J. W. Gregor. 1939. A suggested new terminology. Nature 144:333.

Gilmour, J. S. L. & J. Heslop-Harrison. 1954. The deme terminology and the units of micro-evolutionary change. Genetica 27:147-161.

Gittins, R. 1979. Ecological applications of canonical analysis. Pages 309-535 in L. Orlóci, C. R. Rao & W. M. Stiteler [eds.], Multivariate Methods in Ecological Work. International Co-operative Publishing House, Fairland, Md.

Gladkova, V. N. 1968. Karyological studies of the genera Crataegus L. and Cotoneaster Medic. (Maloideae) as related to their taxonomy. Bot. Žurn. (Moscow & Leningrad) 53:1203-1273.

Gnanadesikan, R. 1977. Methods for Statistical Analysis of Multivariate Observations. John Wiley & Sons, Inc., New York.

Goebel, K. 1900. Organography of Plants. Vol. I. General Organography (Transl. from German by I. B. Balfour). Clarendon Press, Oxford.

Gokhale, V. 1979. Analysis of ecological frequency data: certain case studies. Pages 21-53 in L. Orlóci, C. R. Rao & W. M. Stiteler [eds.], Multivariate Methods in Ecological Work. International Co-operative Publisning House, Fairland, Md.

Goodman, M. N. 1968. A measure of "overall variability" in populations. Biometrics 24:189-192.

Gostyńska-Jakuszewska, M. 1975. The use of numerical methods in the systematics of hawthorns occurring in Poland. Arbor. Kórnickie 20:113-129.

Gottlieb, L. D. 1977. Electrophoretic evidence and plant systematics. Ann. Missouri Bot. Gard. 64:161-180.

Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325-338.

Gower, J. C. 1971a. A general coefficient of similarity and some of its properties. Biometrics 27:857-871.

Gower, J. C. 1971b. Statistical methods of comparing different multivariate analyses of the same data. Pages 138-149 in F. R. Hodson, D. B. Kendall & P. Tautu [eds.], Mathematics in the Archaeological and Historical Sciences. University Press, Edinburgh.

Grant, V. 1975. Genetics of Flowering Plants. Columbia University Press, New York.

Green, P. E. 1978. Analyzing Multivariate Data. Dryden Press, Hinsdale, Ill.

Green, R. H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. John Wiley & Sons, Inc., New York.

Gregor, J. W. & J. M. S. Lang. 1950. Intra-colonial variation in plant size and habit in sea plantains. New Phytol. 49:135-141.

Greyson, R. I., Walden, D. B. & W. J. Smith. 1982. Leaf and stem heteroblasty in Zea. Bot. Gaz. (Crawfordsville) 143:73-78.

Gronovius, J. F. 1739. Flora Virginica, Part 1. Cornelius Haak, Leiden.

Gustafsson, A. 1946. Apomixis in higher plants. I. The mechanism of apomixis. Acta Univ. Lund. n.s. 42(3).

Gustafsson, A. 1947a. Ibid. II. The causal aspect of apomixis. Acta Univ. Lund. n.s. 43(2).

Gustafsson, A. 1947b. Ibid. III. Biotype and species formation. Acta Univ. Lund. n.s. 43(12).

Hall, I. V., Wood, G. W. & L. P. Jackson. 1978. The Biology of Canadian Weeds. 30. Pyrus melanocarpa (Michx.) Willd. Canad. J. Plant Sci. 58:499-504.

Handford, P. 1980. Heterozygosity at enzyme loci and morphological variation. Nature 286:261-262.

Hardin, J. W. 1973. The enigmatic chokecherries (Aronia, Rosaceae). Bull. Torrey Bot. Club 100:178-184.

Harlan, J. R. & J. M. J. deWet. 1975. On O. Winge and a prayer: the origins of polyploidy. Bot. Rev. (Lancaster) 41:361-390.

Harris, R. J. 1975. A Primer of Multivariate Statistics.

Academic Press, New York.

Haskell, G. 1966. The history, taxonomy and breeding system of apomictic British Rubi. Pages 141-151 in J. G. Hawkes [ed.], Reproductive Biology and Taxonomy of Vascular Plants. Pergamon Press, New York.

Heck, D. L. 1960. Charts of some upper percentage points of the distribution of the largest characteristic root. Ann. Math. Stat. 31:625-642.

Herr, M. J. 1971. A new clearing-squash technique for the study of ovule development in Angiosperms. Amer. J. Bot. 58:785-790.

Heslop-Harrison, J. & Y. Heslop-Harrison. 1970. Evaluation of pollen viability by enzymatically induced fluorescence; intracellular hydrolysis of fluorescein diacetate. Stain Technol. 45:115-120.

Heslop-Harrison, Y. & K. R. Shivanna. 1977. The receptive surface of the Angiosperm stigma. Ann. Bot. (London) 41:1233-1258.

Hickey, L. J. 1973. Classification of the architecture of dicotyledonous leaves. Amer. J. Bot. 60:17-33.

Hill, R. S. 1980. A stopping rule for partitioning dendrograms. Bot. Gaz. (Crawfordsville) 141:321-324.

Hoover, K. B. 1961. Hawthorn studies in Pennsylvania in relation to wildlife values. Special Report No. 5, Pennsylvania Cooperative Wildlife Research Unit, University Park, Pa.

Hufford, L. D. 1980. Staminal vascular arcnitecture in five dicotyledonous Angiosperms. Proc. Iowa Acad. Sci. 87:96-102.

Jain, S. K. 1959. Male sterility in flowering plants. Bibliographia Genetica 18:101-166.

Jain, S. K. & P. L. Workman. 1967. Generalized F statistics and the theory of inbreeding and selection. Nature 214:674-678.

Jardine, N. 1971. Patterns of differentiation between human local populations. Philos. Trans. Roy. Soc. B. 263:1-33.

Jardine, N. & R. Sibson. 1970. Quantitative attributes in

taxonomic descriptions. Taxon 19:862-870.

Jardine, N. & R. Sibson. 1971. Mathematical Taxonomy. John Wiley & Sons, Ltd., London.

Jones, G. N. 1946. American species of Amelanchier. Illinois Biol. Monogr. 20:1-126.

Kalkman, K. 1973. The Malesian species of the subfamily Maloideae (Rosaceae). Blumea 21:413-442.

Kania, W. 1973. Studies in the floral development of Rosaceae. Bot. Jahrb. Syst. 93:175-246.

Kim, J.-O. & F. J. Kohout. 1975. Special topics in general linear models. Pages 368-397 in N. H. Nie, C. H. Hull, J. G. Jenkins, K. Steinbrenner & D. H. Bent [eds.], SPSS, Statistical Package for the Social Sciences, 2nd Ed. McGraw-Hill Book Co., New York.

Knuth, P. 1906. Handbook of Flower Pollination, Vol. II (Trans. from German by J. R.Ainsworth Davis). Clarendon Press, Oxford

Koehne, E. 1890. Die Gattungen der Pomaceen. Wissenschaftliche Beilage zum Programm des Falk-Realgymnasiums. Ostern 1890. Programm Nr. 95. R. Gaertner, Berlin.

Kovanda, M. 1961. Flower and fruit morphology of Sorbus in correlation to the taxonomy of the genus. Preslia (Praha) 33:1-16.

Kruschke, E. P. 1955. The Hawthorns of Wisconsin, Part I. Status, objectives, and methods of collecting and preparing specimens. Milwaukee Public Mus. Publ. Bot. No. 2.

Kruschke, E. P. 1965. Contributions to the taxonomy of Crataegus. Milwaukee Public Mus. Publ. Bot. No. 3.

Lance G. N. & W. T. Williams. 1967. A general theory of classificatory sorting strategies. I. Hierarchical systems. Comput. J. 9:373-380.

Layard, M. W. J. 1974. A Monte Carlo comparison of tests of equality of covariance matrices. Biometrika 61:461-465.

Lecoq, H. & M. Lamotte. 1847. Catalogue raisonè des Plantes Vasculaires du Plateau Central de la France. Victor Masson, Paris.

Lefkovitch, L. P. 1978. Consensus coordinates from qualitative and quantitative attributes. Biom J. 20:679-691.

Lefkovitch, L. P. 1982. A non-parametric method for comparing dissimilarity matrices (abstract). 16th Numerical Taxonomy Conference, 22-24 October 1982.

Leins, P., Merxmüller, H. & R. Sattler. 1972. Terminology of intercalary cup formations in flowers. Ber. Deutsch. Bot. Ges. 85:294.

Levin, D. A. 1975. Pest pressure and recombination systems in plants. Amer. Nat. 109:437-451.

Lewis, W. H. 1980. Polyploidy in species populations. Pages 103-145 in W. H. Lewis [ed.], Polyploidy - Biological Relevance. Plenum Press, New York & London.

Lewontin, R. C. 1966. On the measurement of relative variability. Syst. Zool. 15:141-142.

Liljefors, A. 1953. Studies of propagation, embryology and pollination in Sorbus. Acta Horti Berg. 16:277-329.

Liljefors, A. 1954. Cytological studies in Sorbus. Acta Horti Berg. 17:47-113.

Lindley, J. 1822. Observations on the natural group of plants called Pomaceae. Trans. Linn. Soc. London 13:88-106.

Linnaeus, C. 1753. Species Plantarum, Vol. 1. L. Salvius, Stockholm. [microfiche edition, International Documentation Centre AB, Tumba, Sweden]

Lloyd, D. G. Benefits and handicaps of sexual reproduction. Pages 69-111 in M. K. Hecht, W. C. Steere & B. Wallace [eds.], Evolutionary Biology, Vol. 13. Plenum Press, New York.

Longley, A. E. 1924. Cytological studies in the genus Crataegus. Amer. J. Bot. 11:295-317.

Loudon, J. C. 1838. Arboretum et Fruticetum Brittanicum. Longman, Orme, Brown, Greene & Longmans, London.

Love, R. & M. Feigen. 1978. Interspecific hybridization between native and naturalized Crataegus in western Oregon. Madroño 25:211-217.

Luckwill, L. C. 1970. The control of growth and fruitfulness in apple trees. Pages 237-253 in L. C. Luckwill & C. V. Cuttings [eds.], Physiology of Tree Crops. Academic Press, London.

Macleod, M. 1980. Pollen methods and analysis. Unpublished report.

Mahalanobis, P. C. 1936. On the generalized distance in statistics. Proc. Natl Inst. Sci., India 2:49-55.

Marsnall, D. R. & A. H. D. Brown. 1981. The evolution of apomixis. Heredity 47:1-15.

Marshall, D. R. & B. S. Weir. 1979. Maintenance of genetic variation in apomictic plant populations. I. Single locus models. Heredity 42:159-172.

Marie-Victorin, Frère. 1938. Phytogeographical problems of eastern Canada. Amer. Midl Naturalist 19:489-558.

Martin, A. C., Zim, A. S. & A. L. Nelson. American Wildlife and Plants - A Guide to Wildlife Food Habits. McGraw-Hill Book Co., New York.

McAllister, H. A. & C. M. Gillham. 1980. Tab. 792 Sorbus forrestii Rosaceae Bot. Mag. 83:1-4.

McNeill, J. The handling of character variation in numerical taxonomy. Taxon 23:699-705.

Meltzer, B., Searle, N. H. & R. Brown. 1967. Numerical specification of biological form. Nature 216:32-37.

Melville, R. 1953. Growth and plant systematics. Proc. Linn. Soc. London 164:173-181.

Mitton, J. B. 1978. Relationship between heterozygosity for enzyme loci and variation in morphological characters in natural populations. Nature 273:661-662.

Moffett, A. A. 1931. A preliminary account of chromosome behavior in the Pomoideae. J. Pomol. Hort. Sci. 9:100-110.

Mosimann, J. E. & F. C. James. 1979. New statistical methods for allometry with application to Florida Red-winged Blackbirds. Evolution 33:444-459.

Muniyamma, M. & J. B. Phipps. 1979a. Cytological proof of apomixis in Crataegus L. Amer. J. Bot. 66:149-155.

Muniyamma, M. & J. B. Phipps. 1979b. Studies of pollen meiosis and polyploidy with relation to Crataegus systematics in Ontario. Canad. J. Genet. Cytol. 21:231-241.

Muniyamma, M. & J. B. Phipps. in press. Agamospermy in Crataegus. J. Linn. Soc., Bot.

Nettancourt, D. de 1977. Incompatibility in Angiosperms. Springer-Verlag, Berlin.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. & D. H. Bent. 1975. SPSS, Statistical Package for the Social Sciences, 2nd ed. McGraw-Hill Book Co., New York.

Nordenskiöld, E. 1928. The History of Biology (Trans. from Swedish by L. B. Eyre). Alfred A. Knopf, Inc., New York.

Nuñez, O. 1968. An acetic haematoxylin squash method for small chromosomes. Caryologia 21:115-119.

Nygren, A. 1967. Apomixis in Angiosperms. Pages 551-596 in W. Ruhland [ed.], Encyclopedia of Plant Physiology, Vol. 18. Springer Verlag, Berlin.

Orlóci, L. 1968. A model for the analysis of structure in taxonomic collections. Canad. J. Bot. 46:1093-1097.

Orlóci, L. 1973. Ranking characters by a dispersion criterion. Nature 244:371-373.

Orlóci, L. 1975. Measurement of redundancy in species collections. Vegetatio 31:65-67.

Orlóci, L. 1978. Multivariate Analysis in Vegetation Research, 2nd Ed. Dr.. W. Junk B. V., The Hague.

Orlóci, L. & M. E. Beshir. 1976. A heuristic test for homogeneity in species populations. Vegetatio 31:141-145.

Oxnard, C. E. 1978. One biologist's view of morphometrics. Ann. Rev. Ecol. Syst. 9:219-241.

Palmer, E. J. 1925. Synopsis of North American Crataegi. J. Arnold Arbor. 6:5-128.

Palmer, E. J. 1932. The Crataegus problem. J. Arn. Arb. 13:342-362.

Palmer, E. J. 1946. Crataegus in the northeastern and

central United States and adjacent Canada. Brittonia 5:471-490.

Palmer, E. J. 1962. Crataegus. Pages 148-164 in H. Kurz & R. K. Godfrey, Trees of Northern Florida. University of Florida Press, Gainesville.

Palmer, E. J. 1963. Crataegus. Pages 338-375 in H. A. Gleason & A. Cronquist, The New Britton and Brown Illustrated Flora of the Northeastern United States and Adjacent Canada, Vol. 2. Hafner Publishing Co., for the New York Botanical Garden, New York.

Payer, J.-B. 1857. Traitè d'Organogènie comparèe de la Fleur. Victor Masson, Paris.

Pearson, K. 1926. On the coefficient of racial likeness. Biometrika 18:105-117.

Pèchoutre, F. 1902. Contribution à l'étude du dèveloppement de l'ovule et de la graine des Rosacées. Ann. Sci. Nat. Bot. Sèr. 8, 16:1-158.

Phipps, J. B. 1970. Studies in the Arundinelleae (Gramineae). X. Preliminary taximetrics. Canad. J. Bot. 48:2333-2356.

Phipps, J. B. 1971. Dendrogram topology. Syst. Zool. 20:306-308.

Phipps, J. B. in press, a. Crataegus - a nomenclator for sectional or serial names. Taxon.

Phipps, J. B. in press, b. Biogeographic, taxonomic and cladistic relationships between east Asiatic and Nortn American Crataegus. Ann. Missouri Bot. Gard.

Phipps, J. B. & M. Muniyamma. 1980. A taxonomic review of Crataegus (Rosaceae) in Ontario. Canad. J. Bot. 58:1621-1699.

Pillai, K. C. S. 1960. Statistical Tables for Tests of Multivariate Hypotheses. Statistical Center, University of the Philippines, Manila.

Pimentel, R. 1979. Morphometrics - The Multivariate Analysis of Biological Data. Kendall/Hunt, Dubuque, Iowa.

Plukenet, L. 1692. Phytographia, Part III. By the author, London.

Podani, J. 1980. SYN-TAX: Computer program package for cluster analysis in ecology, phytosociology and taxonomy. Abstracta Botanica (Budapest), 6.

Podani, J. 1982. Spatial processes in the analysis of vegetation. Ph.D. Thesis. University of Western Ontario.

Posluszny, U., Scott, M. G. & R. Sattler. 1980. Revisions in the technique of epi-illumination light microscopy for the study of floral and vegetative apices. Canad. J. Bot. 58:2491-2495.

Power, N. 1980. Crataegus-insect relationships. Unpublished report.

Prentice, H. C. 1979. Numerical analysis of infraspecific variation in European Silene alba and S. dioica (Caryophyllaceae). J. Linn. Soc., Bot. 78:181-212.

Proctor, M. & P. Yeo. 1973. The Pollination of Flowers. Collins, London.

Provancher, L. 1862. Flore Canadienne. Joseph Darveau, Québec.

Rao, C. R. 1952. Advanced Statistical Methods in Biometric Research. John Wiley & Sons, Inc., New York.

Raup, H. M. 1966. The view from John Sanderson's farm. Forest History 10:1-11.

Reader, R. J. 1975. Effect of air temperature on the flowering date of dogwood (Cornus florida). Canad. J. Bot. 53:1523-1534.

Reyment, R. A. 1971. Multivariate normality in morphometric analysis. Math. Geol. 4:357-368.

Richards, A. J. 1975. Sorbus L. Pages 233-243 in C. A. Stace [ed.], Hybridization and the Flora of the British Isles. Academic Press, London.

Rickett, H. W. 1936. Forms of Crataegus pruinosa. Bot. Gaz. (Crawfordsville) 97:788-793.

Rickett, H. W. 1937. Forms of Crataegus crus-galli. Bot. Gaz. (Crawfordsville) 98:609-616.

Rickett, H. W. 1943. The inflorescence of Crataegus. Bull. Torrey Bot. Club 70:489-495.

Rieger, R., Michaelis, A. & M. M. Green. 1968. A Glossary of Genetics and Cytogenetics, 3rd Ed. Springer-Verlag, Berlin.

Robertson, K. R. 1974. The genera of the Rosaceae in the southeastern United States (Rosoideae, Maloideae, Amygdaloideae). J. Arnold Arbor. 55:611-662.

Robinson, W. A. 1982. Experimental taxonomy in the genus Amelanchier. II. Do the taxa in the genus Amelanchier form an agamic complex? Rhodora 84:85-100.

Robinson, W. A. & C. R. Partanen. 1980. Ibid. I. A new look at the chromosome numbers of the Amelanchier species growing in the northeastern United States. Rhodora 82:483-493.

Rohlf, F. J. 1977. A note on the measurement of redundancy. Vegetatio 34:63-64.

Rohlf, F. J. & R. R. Sokal. 1969. Statistical Tables. W. H. Freeman & Co., San Francisco.

Rohlf, F. J. & R. R. Sokal. 1981. Comparing numerical taxonomic studies. Syst. Zool. 30:459-490.

Ryan, T. A. & B. L. Joiner. Undated. Normal probability plots and tests for normality. Technical Report, Statistics Department, The Pennsylvania State University, University Park, Pa.

Ryan, T. A., Joiner, B. L. & B. F. Ryan. 1976. MINITAB Student Handbook. Duxbury Press, North Scituate, Mass.

Ryan, T. A., Joiner, B. L. & B. F. Ryan. 1981. Minitab Reference Manual. Minitab Project, The Pennsylvania State University, University Park, Pa.

Sanghvi, L. D. 1953. Comparison of genetical and morphological methods for a study of biological differences. Amer. J. Phys. Anthropol. 11:385-404.

Sargent, C. S. 1892. Rosaceae to Saxifragaceae. Silva of North America, Vol. 4. Riverside Press, Cambridge, Mass.

Sargent, C. S. 1902. Rhamnaceae to Rosaceae. Silva of North America, Supplement Vol. 13. Riverside Press, Cambridge, Mass.

Sargent, C. S. 1902. Crataegus pushii Bot. Gaz. (Crawfordsville) 33:109-110.

Sargent, C. S. 1907. The genus Crategus in North America. J. Bot. 45:289-292.

Sargent, C. S. 1908. Crataegus in Ontario. Ontario Nat. Sci. Bull. 1908:11-98.

Sarvas, R. 1967. Climatological control of flowering in Trees. Proceedings, 14th IUFRO Congr. 3(22):15-30.

Sass, J. E. 1958. Botanical Microtechnique, 3rd Ed. Iowa State University Press, Ames, Iowa.

Sattler, R. 1968. A technique for the study of floral development. Canad. J. Bot. 46:720-722.

Sax, H. J. 1954. Polyploidy and apomixis in Cotoneaster. J. Arnold Arbor. 35:334-365.

Schaal, B. A. 1982. Variation in nuclear DNA sequences of plant populations (abstract). Bot. Soc. Amer. Misc. Publ. No. 162.

Schneider, C. K. 1906. Illustrierte Handbuch der Laubholzkunde. Verlag Gustav Fischer, Jena.

Seal, H. L. 1964. Multivariate Statistical Analysis for Biologists. Methuen & Co., Ltd., London.

Sharma, A. K. & A. Sharma. 1965. Chromosome Techniques, Theory and Practice. Butterworths, Washington, D. C.

Sibson, R. 1969. Information radius. Z. Wahrscheinlichskeittheorie u. verw. Geb. 14:149-160.

Simpson, G. G., Roe, A. & R. C. Lewontin. 1960. Quantitative Zoology, Rev. Ed. Harcourt, Brace & Co., New York.

Sinnott, Q. P. 1978. A taxonomic study of series Pruinosae (Crataegus, Rosaceae). M.Sc. Thesis. University of Western Ontario.

Sinnott, Q. P. 1980. A simple similarity coefficient for taxonomic comparisons. Taxon 30:18-26.

Sinnott, Q. P. & J. B. Phipps. 1983. Variation patterns in Crataegus Series Pruinosae (Rosaceae) in southern Ontario. Syst. Bot. 8:59-70.

Small, E., Lefkovitch, L. P. & D. Classen. 1982. Character set incongruence in Medicago. Canad. J. Bot.

60:2505-2510.

Smith, G. L. 1963a. Studies in Potentilla L. I. Embryological investigations into the mechanism of agamospermy in British P. tabernaemontani Aschers. New Phytol. 62:264-282

Smith, G. L. 1963b. Ibid. II. Cytological aspects of apomixis in P. crantzii (Cr.) Beck ex Fritsch. New Phytol. 62:283-300.

Smith, G. L. 1971. Ibid. III. Variation in British P. tabernaemontani Aschers and P. crantzii (Cr.) Beck ex Fritsch. New Phytol. 70:607-618.

Smith, J. M. 1978. The Evolution of Sex. Cambridge University Press, Cambridge.

Smith, P. G., Phipps, J. B. & T. A. Dickinson. 1980. Accumulated heat in relation to Crataegus flowering (abstract). Contributed paper, 2nd Intl. Congr. Syst. Evol. Biol.

Sneath, P. H. A. 1977. A method for testing the distinctness of clusters: a test of the disjunction of two clusters in euclidean space as measured by their overlap. Math. Geol. 9:123-143.

Sneath, P. H. A. 1979a. BASIC program for a significance test for clusters in UPGMA dendrograms obtained from squared euclidean distances. Computers and Geosciences 5:127-137.

Sneath, P. H. A. 1979b. BASIC program for a significance test for two clusters in euclidean space as measured by their overlap. Computers and Geosciences 5:143-155.

Sneath, P. H. A. 1979c. BASIC program for nonparametric significance of overlap between a pair of clusters using the Kolmogorov - Smirnov test. Computers and Geosciences 5:173-188.

Sneath, P. H. A. & R. R. Sokal. 1973. Numerical Taxonomy. W. H. Freeman & Co., San Francisco.

Sokal, R. R. & C. A. Braumann. 1980. Significance tests for coefficients of variation and variability profiles. Syst. Zool. 29:50-66.

Sokal, R. R. & F. J. Rohlf. 1962. The comparison of dendrograms by objective methods. Taxon 11:33-40.

Sokal, R. R. & F. J. Rohlf. 1969. Biometry. W. H. Freeman & Co., San Francisco.

Sokal, R. R. & F. J. Rohlf. 1981. Biometry, 2nd ed. W. H. Freeman & Co., San Francisco.

Standish, L. M. 1916. What is happening to the hawthorns? J. Heredity 7:266-279.

Stanley, R. G. & H. F. Linskens. 1974. Pollen: Biology, Biochemistry and Management. Springer-Verlag, Berlin.

Stebbins, G. L., Jr. 1941. Apomixis in the Angiosperms. Bot. Rev. (Lancaster) 7:507-542.

Stebbins, G. L., Jr. 1950. Variation and Evolution in Plants. Columbia University Press, New York.

Sterling, C. 1964. Comparative morphology of the carpel in the Rosaceae, III. Pomoideae: Crataegus, Hesperomeles, Mespilus, Osteomeles. Amer. J. Bot. 51:705-712.

Strahler, A. H. 1978. Binary discriminant analysis: a new method for investigating species-environment relationships. Ecology 59:108-116.

Sturges, H. A. 1926. The choice of a class interval. J. Amer. Statist. Ass. 21:65-66.

Sutton, S. B. 1970. Charles Sprague Sargent and the Arnold Arboretum. Harvard University Press, Cambridge, Mass.

Synnott, D. M. 1978. The status of Crataegus laevigata (Poiret) DC. in Ireland. Glasra 2:49-55.

Thiselton-Dyer, W. T. 1904. Index Kewensis, Supplement Vol. 2, 1896-1900. Clarendon Press, Oxford.

Thorndike, R. M. 1978. Correlational Procedures for Research. Gardner Press, New York.

Usberti, J. A., Jr. & S. K. Jain. 1978. Variation in Panicum maximum: a comparison of sexual and asexual populations. Bot. Gaz. (Crawfordsville) 139:112-116.

Valek, D. 1980. Hawthorn Type. Page 38 in F. H. Eyre [ed.], Forest Cover Types of the United States and Canada. Society of American Foresters, Washington, D. C.

Van Valen, L. 1974. Multivariate structural statistics in natural history. J. Theor. Biol. 45:235-247.

Van Valen, L. 1976. Ecological species, multispecies and Oaks. Taxon 25:233-239.

Van Valen, L. 1978. The statistics of variation. Evolutionary Theory 4:33-43.

Wainio, W. W. & E. B. Brooks. 1941. The chemical composition of forest fruits and nuts from Pennsylvania. J. Agric. Res. 62:527-635.

Walbot, V., Beachy, R. N. & M.-C. Yao. 1980. Molecular techniques applied to polyploids. Pages 529-537 in W. H. Lewis [ed.], Polyploidy - Biological Relevance. Plenum Press, New York.

Warwick, S. I. & J. McNeill. 1982. Population variation in Plantago major from southern Ontario. Canad. J. Bot. 60:2846-2856.

Webber, L. R. & D. W. Hoffmann. 1967. Origin, Classification and Use of Ontario Soils. Ontario Department of Agriculture and Food, Toronto.

Wellhouse, W. H. 1922. The insect fauna of Crataegus. Memoir 56, Cornell University Agricultural Experiment Station.

Westwood, M. N. & H. I. Bjornstad. 1971. Some fruit characteristics of interspecific hybrids and extent of self-sterility in Pyrus. Bull. Torrey Bot. Club 98:22-24.

Whitney, G. G. 1982. Vegetation-site relationships in the presettlement forests of northeastern Ohio. Bot. Gaz. (Crawfordsville) 143:225-237.

Więckowska, I. 1970. Studies on the morphogenesis of tree leaves. I. The shape of leaves on short shoots of Ulmus laevis Pall. Acta Soc. Bot. Poloniae 39:751-768.

Wilks, S. S. 1932. Certain generalizations in the analysis of variance. Biometrika 24:471-474.

Williams, G. C. 1975. Sex and Evolution. Monographs in Population Biology No. 8, R. M. May, ed. Princeton University Press, Princeton.

Williams, R. R. 1970. Factors affecting pollination in fruit trees. Pages 193-207 in L. C. Luckwill & C. V. Cutting [eds.], Physiology of Tree Crops. Academic Press, London.

Williams, W. T. & H. T. Clifford. 1971. On the comparison of two classifications of the same set of elements. Taxon 20:519-522.

Wishart, D. 1978. CLUSTAN User Manual, 3rd Ed. Rep. No. 47, Inter-University/Research Councils Series. Program Library Unit, Edinburgh University.

Yablokov, A. V. 1974. Variability in Mammals. Amerind, New Delhi.

END

20030184

FIN