Morphometric discrimination among sibling species in the *fulva – rudis – texana* complex of the ant genus *Aphaenogaster* (Hymenoptera: Formicidae)

Gary J. Umphrey

Abstract: The *fulva – rudis – texana* complex of the ant genus *Aphaenogaster* includes *A. fulva*, *A. rudis*, *A. texana*, and morphologically similar species. Morphometric and other morphological investigations were conducted on colony representatives of 10 forms (4 likely representing undescribed species) that were previously identified using cytogenetic and electrophoretic markers. In workers, most qualitative characters exhibit such a high degree of size-associated intraspecific variation relative to interspecific differences that they are not reliable for identification. Linear discriminant analysis and canonical variate analysis on 12 morphometric measurements were used to optimize classification and define a morphometric habitus for each form. Two-variable scatterplots clarify the nature of morphometric variation in the complex and provide simple characters that will reliably separate numerous pairs of forms. Indices appear to be of little taxonomic value in this complex. A preliminary key to the workers of the complex summarizes the most important taxonomic characters. This key substantially improves the ability to recognize morphologically most members of the complex, but sometimes only genetic evidence is definitive.

Résumé : Chez les fourmis du genre *Aphaenogaster*, le complexe *fulva – rudis – texana* se compose d'*A. fulva*, *A. rudis*, *A. texana* et des espèces morphologiquement semblables. Des examens morphométriques et morphologiques ont été effectués chez des représentants de colonies de 10 formes différentes (dont 4 probablement d'espèces non décrites) préalablement reconnues au moyen de marqueurs cytogénétiques et électrophorésiques. Chez les ouvrières, le plupart des caractéristiques qualitatives subissent un tel degré de variation intraspécifique associée à la taille comparativement aux différences interspécifiques, qu'elles ne peuvent servir à l'identification. Une analyse discriminante linéaire et une analyse des variables canoniques de 12 mesures morphométriques ont permis d'optimiser la classification et de définir un habitus morphométrique pour chaque forme. Des diagrammes de dispersion impliquant deux variables précisent la nature de la variation morphométrique au sein du complexe et permettent d'identifier des caractères simples qui distinguent très bien plusieurs paires de formes. Les coefficients ont une valeur taxonomique douteuse pour identifier les fourmis de ce complexe. Une clé préliminaire des ouvrières de ce complexe résume les principales caractéristiques diagnostiques. Cette clé permettra de reconnaître plus facilement la plupart des membres du complexe par des caractères morphologiques, mais parfois seuls les caractères génétiques garantissent une identification exacte.
[Traduit par la Rédaction]

Introduction

This paper examines morphological differentiation, largely through morphometric methods, in the "*fulva – rudis – texana*" complex; a quick definition of this complex is that a member worker will reach couplet 18 in Creighton's (1950) key to the North American *Aphaenogaster*. Although these are some of the most abundant ants in the woods of eastern North America, their taxonomy requires revision. Previous work (Carroll 1975; Crozier 1975, 1977; personal observation) suggested that standard morphological procedures alone would be inadequate to resolve the taxonomy of this complex, but Crozier showed that karyology and electrophoresis would be highly informative. My approach attempts to definitively resolve the tangled systematics of this complex by integrating these genetic methods with morphological analysis.

Received June 23, 1995. Accepted November 24, 1995.

G.J. Umphrey. Department of Zoology, The University of Western Ontario, London, ON N6A 5B7, Canada (e-mail: umphrey@julian.uwo.ca).

On the basis of cytogenetic and electrophoretic markers, Umphrey (1992; G.J. Umphrey, unpublished data) recognized 10 genetically differentiated forms among 223 colonies from 63 localities collected primarily throughout eastern North America (but also Arizona, Missouri, and Texas; see Table 2). I believe that these 10 forms represent *A. fulva* and a *rudis – texana* subcomplex of nine sibling species. Since better evidence of reproductive isolation is desirable in a couple of cases, I usually use "form" in place of "species" throughout this paper, but I expect these terms to prove to be synonymous. While basic questions of species status and nomenclature are unravelled, the nine forms in the *rudis – texana* subcomplex are labelled using a code that incorporates the chromosome number of the haploid karyotype (excluding supernumeraries). Table 1 summarizes the codes and the names used (as species or subspecies) by Creighton (1950) that I believe apply to each form; the name *silvestrii*, described by Menozzi (1929) and incorrectly synonymized with *texana* by Creighton (1950), may be a senior synonym of *miamiana*. Umphrey (1992; G.J. Umphrey, unpublished data) gives a genetic diagnosis and distribution map for each form. (The rare desert species *A. punctaticeps* can also be

included in this complex, but it is treated only briefly here; further analysis will be provided by S.P. Cover and G.J. Umphrey (in preparation).)

Specimens from genetically identified colonies allow morphological differentiation to be studied with greater confidence in the initial delimitation of the species. Morphometric analysis was motivated by the apparent scarcity of simple external characters of taxonomic value. Discriminant analysis and canonical variate analysis attempted to optimize classification using 12 morphometric variables, and to compare the taxonomic value of these variables. These multivariate methods also provide a "morphometric habitus" for each form. Numerous two-variable scatterplots are used to simplify and further illustrate the nature of intraspecific and interspecific morphometric variation. Seven scatterplots provide an empirical test of the taxonomic value of indices for this complex.

Rationale for a morphometric approach

When individuals can be reliably identified to group on the basis of cryptic characters (e.g., genetic markers, genitalia), these samples can be used to search for external morphological characters that can be used to classify individuals for which the cryptic characters are either unavailable or available at greater cost. When several measurement characters are used, discriminant analysis and canonical variate analysis are powerful multivariate methods of examining and summarizing the data. Discriminant analysis constructs functions that minimize classification errors for a given criterion (e.g., overall misclassification rate, costs of misclassification; see Lachenbruch 1975, 1982). By using a variable selection procedure (e.g., stepwise), weak or redundant variables can be eliminated from the discriminant model and the relative value of the useful variables in the model can be adduced. Linear discriminant analysis assumes that the populations have multivariate normal densities and a common covariance matrix. Lubischew (1962) demonstrated how the choice of characters affects discriminant ability, using examples from sibling species complexes in two flea beetle genera. Canonical variate analysis provides a powerful method of reducing the dimensionality of the data by constructing new canonical variables, uncorrelated with one another, that are linear functions of the original variables (see Blackith and Reyment 1971).

Multivariate methods have not been used much in ant systematics, although several studies (e.g., Francoeur 1973; Ward 1985, 1989, 1993; Wing 1968) have made extensive use of measurement data and scatterplots of pairs of variables (including ratio variables). Principal component analysis was used by Cagniant (1989) to compare morphological variation in three populations of *Aphaenogaster weulersseae*, and also by Crozier et al. (1986) to study evolutionary patterns in some species of Australian *Rhytidoponera*. Reyes Lopez and Porras Castillo (1984) used linear discriminant analysis on 10 forewing measurements to obtain perfect separation of queens from five nests (52 specimens) of *Goniomma hispanicum* and four nests (59 specimens) of *G. baeticum*, but it is never clear in their analysis how reducing variables might affect classification. The general applicability of their analysis is questionable, since the number of colonies used was small relative to the number of variables, and information on the geographical distribution of the sampled colonies is wanting. Elmes (1978) used canonical variate analysis on 12 variables to get excellent morphological separation of the males and queens of three species of *Myrmica*. His keys, based entirely on canonical variate scores, have the unusual feature of giving a confidence level for correct classification.

Table 1. Codes used in this paper for forms in the *fulva–rudis–texana* complex of *Aphaenogaster*, with corresponding names used by Creighton (1950).

Code	Name
N16	Undescribed
N17	Undescribed
N18	*picea*
N19	Undescribed
N20	*carolinensis*
N21a	*miamiana*
N21b	*texana*
N22a	*rudis*
N22b	Undescribed

Materials and methods

Samples of specimens from a colony were collected into 85% ethanol and designated as either the primary or a secondary sample. An *Aphaenogaster* colony number was assigned (A001 to A446 in this study). The primary sample was taken preferably when the colony was first collected. Secondary samples collected in the laboratory could include reared reproductives, nest queens, unusually small or large workers, and abnormal individuals. Long laboratory rearing often produced minim workers, although none were present when the colony was first collected; these small workers (usually the only size class found in incipient colonies) have often caused taxonomic confusion. Any sample not preserved when the colony was first collected was also labelled with the date of preservation.

Workers in the primary sample were laid out in rows, then six were selected using random-number tables, mounted individually, and labelled according to their position in the random sample. This was not a truly random sample from all workers present in the colony, since the primary sample itself could not be obtained randomly, but the greater source of bias likely comes in the laboratory when the specimens are selected (with a probable tendency towards mounting larger ants).

From the remainder, two further pins of three workers (small, medium, large) were usually mounted, as well as selected males and gynes and any anomalous ants. Under normal air-drying, various sclerites of the males tend to collapse. Superior male specimens were prepared by John Swann using the critical-point dryer at the University of Guelph. Studies of morphological differentiation were aided by the preparation of representative scanning electron micrographs of the (*i*) head, (*ii*) alitrunk, and (*iii*) petiole plus postpetiole by Lewis Ling of Carleton University. Male genitalia were soaked in 10% KOH for 3 h at room temperature, rinsed thoroughly with distilled water, and stored in individual genital capsules in a drop of glycerin.

The morphometric analysis in this paper is based on 12 measurements taken on the first randomly selected worker for each colony used. The additional material was used to make qualitative assessments of morphological differentiation and will serve as a base for further research.

Measurements were made at a magnification of 50× (25× if the measurement exceeded 2.00 mm) using a Nikon SMZ2T zoom dissecting microscope fitted with an ocular scale (laboratory of S. Marshall, University of Guelph). The scale was calibrated and checked frequently for accuracy with an ocular micrometer. The 12 measurements made (usually on the left side of the body, but sometimes on the right if the body part was damaged or obscured) are as follows.

HW	Head width	With head in full-face view, measured directly above the eyes
HL	Head length	Maximum length down the centre of the head from the top of the head to the top of the clypeal notch. This measurement differs slightly from HL of most other ant taxonomists
SL	Scape length	Maximum linear measurement of the length of the antennal scape (first antennal segment), excluding the basal condyle
EL	Eye length	Maximum length (diameter) of eye in lateral view
WLA	Weber's length of alitrunk	Similar to WL of Ward (1980), which was defined as a diagonal measurement "from the anterior margin of the pronotum (excluding the collar) to the posterior extremity of the metapleural lobe." The measurement was taken with the specimen in approximate lateral view but slightly tilted so that both reference points were sharply focused
MH	Mesonotal height	Maximum measurement from the promesonotal suture to the posteroventral corner of the katepisternum
SPL	Spinal length	Length of the propodeal spine, measured from the centre of the propodeal spiracle to the tip of the spine
ISPL	Interspinal length	Distance from the centre of a straight line extending between the tips of the propodeal spines and the midpoint on the propodeum between the spines. The specimen was oriented to maximize this measurement. If a spine was broken, its length was estimated from that of the other spine
SPD	Spinal distance	Distance between the tips of the propodeal spines
FW	Femur width	Maximum width (diameter) of the hind femur, measured orthogonal to the line of measurement of FL
FL	Femur length	Maximum length of the hind femur *plus* trochanter. This is easier and more precise than measuring the femur only
TL	Tibia length	Maximum length of the hind tibia

Measurements were made to the nearest 0.01 mm (one-half of the smallest scale unit at 50×). The greatest variability in measurements occurs with MH and WLA, which present difficulties in positioning and defining limits.

Despite the well-known criticisms of the use of ratios and indices in systematic studies as a method of scaling for size (Atchley et al. 1976), indices (or ratios) have a long tradition of use in ant systematics. Hence, the following indices were calculated in order to empirically examine their taxonomic value in this complex: CI (cephalic index) = 100(HW/HL); SI1 (scape index 1) = 100(SL/HW); SI2 (scape index 2) = 100(SL/HL); EI (eye, or ocular, index) = 100(EL/HW); AI (alitrunk index) = 100(MH/WLA); SPI (spinal index) = 100(ISPL/HW); FI1 (femur index 1) = 100(FW/FL); FI2 (femur index 2) = 100(FL/HW).

The "regular" data set, used to derive functions for the discriminant and canonical variate analyses and for the two-variable scatterplots, consists of 12 measurements on each of 264 workers in total for the 10 forms. I consider the identity of each worker to be known with certainty. The colonies of 222 of them were karyotyped. Another 40 are from colonies for which karyotype slides were made but the slides were not examined, since I was confident of the identity of the colony. Two workers of N22b are from colonies that could not be karyotyped, but I find this form generally easy to distinguish by habitus. Sample sizes ranged from 5 for N19 to 67 for N22a (see Tables 2 and 3). The functions derived from the regular data set were used to classify the workers in a further "test" data set of 30 observations. These included a paratype worker for each of "*A. huachucana crinimera*," described by Cole (1953), *A. punctaticeps*, described from two workers by MacKay (1989), and *A. huachucana*. Other selected specimens are discussed in the Results.

Statistical analyses were performed using SAS, version 5.12 at the University of Guelph and version 6.03 at Carleton University. All essential documentation for the procedures used (CANDISC, DISCRIM, REG, STEPDISC) can be found in the SAS/STAT user's guide, release 6.03 edition (SAS Institute Inc. 1988). Linear discriminant analysis minimized overall misclassification rate, using equal priors. SigmaPlot version 4.1 was used to prepare Figs. 6–48.

Results

Morphological analyses: qualitative results

Although workers of most forms have a fairly distinctive habitus, differences are rather hard to describe; the most tractable appear to be slight size and shape differences and differences in colour. I have detected nothing of significance in chaetotaxy differences to date, and the petiole and postpetiole appear to offer little of taxonomic value. A rogues' gallery of representative heads (Figs. 1, 2) and alitrunks (Figs. 3, 4) is presented; compare the given measurement data with Table 3 to determine the relative size of a specimen illustrated for its form. Some apparent differences in head shape and length of the propodeal spines, which are amenable to morphometric analysis, can be readily noted, but these differences are confounded by size effects.

Understanding the nature of morphological differentiation in this complex has been impeded by a failure to appreciate how much size variation can occur within each form, and how morphological characters vary with size. For example, the primary component of "head shape" is often expressed by CI (CI = 100HW/HL). In all members of this complex, CI increases, on average, as body size increases (Fig. 42); combined with wide size variation in each form, a smaller "broad-headed" form can have a relatively narrower head (lower CI) than a larger "narrow-headed" form.

As for shape, sculptural differences are highly size

Table 2. Colony composition of the regular data set used in the morphometric analysis of the *fulva–rudis–texana* complex of *Aphaenogaster*.

fulva

FL2: A240. FL4: A094. FL5: A244. GA2: A308–A309, A310*. GA4: A303, A305. IN2: A383*, A385*. KY1: A339*. MD1: A005–A006. MO2: A378*. NC1: A277. NC4: A122. NC7: A285. NC8: A264*. NC11: A126. NJ1: A170, A173. NJ3: A167–A168. NJ4: A151, A154, A158–A159. OH1: A394*. SC1: A102. SC2: A292–A294. SC4: A105–A107, A109, A112. TN2: A338*.

N16

NJ2: A185. OH2: A404. ON1: A201–A202, A424, A427, A429. ON2: A430. ON5: A188, A230–A231, A409, A410*. ON6: A058. PA2: A361.

N17

KY2: A345–A347, A349–A350. OH2: A400, A401*, A402. ON3: A228. ON5: A187, A189–A191, A193–A194, A204, A207–A210, A212, A235–A237, A411*. ON6: A412–A414, A416–A420.

N18

CT1: A251*, A252–A254, A255*. GA3. A318–A321, A323–A325. NC3: A330, A331*, A332. NC10: A250. NY1: A366*. ON1: A203, A423, A426, A428. ON3: A223, A225, A229. ON4: A150. PA1: A405*, A406–A407, A408*. PA4: A362*, A363, A364–A365*. PA5: A029. WV1: A353. WV2: A355–A356, A357–A358*.

N19

MO1: A368, A372. MO2: A375–A377.

N20

FL5: A246. GA1: A311–A315. NC2: A281. NC5: A274, A275*. NC6: A290–A291. NC7: A284, A287*. NC8: A266, A270–A273. NJ1: A177–A182. SC1: A103, A296–A300. SC2: A295. SC3: A117.

N21a

FL1: A247–A249. FL2: A239, A241–A243. FL3: A096–A097. FL6: A238. GA4: A304, A306–A307.

N21b

AZ1: A432–A438. AZ2: A439–A443. TX1: A422. (Counts were verified for all colonies, but only A422, A435–A436, and A441–A442 were examined with Carleton University's Zeiss microscope.)

N22a

AL1: A085. IN1: A382. IN2: A384, A386–A388*. IN3: A389, A391–A392*. KY1: A073, A341–A344. KY2: A348, A351–A352. MD1: A009. MD2: A359*. MO1: A369–A370, A373. NC1: A278. NC2: A279–A280, A282–A283. NC4: A121, A123. NC9: A326–A328. NC11: A125. NJ1: A004, A169, A172, A174–A175. NJ2: A160–A166. NJ4: A155–A157. NJ5: A183–A184, A186. OH1: A393, A395–A397. PA3: A017, A019. TN1: A080. TN2: A334–A337. VA1: A256–A258, A259–A260*.

N22b

IN3: A390. MD2: A360. M01: A367†. ON6: A218, A219†, A220, A222, A415, A421.

Locality codes: CANADA: Ontario: ON1, Chaffey's Locks; ON2, Constance Bay (30 km W Ottawa); ON3, 15 km E Gananoque; ON4, 6 km W Halls Lake, near Lake Kushog; ON5, Pinery Provincial Park (4 km S Grand Bend); ON6, Rondeau Provincial Park (15 km S Ridgetown). U.S.A.: Alabama: AL1, Cleburne Co., Cheaha State Park, 2300 ft. Arizona: AZ1, Cochise Co., 13 km WNW Portal, Pinery Canyon, Chiricahua Mountains, 6200 ft; AZ2, Pima Co., 16 km NE Tucson, Bear Canyon, Santa Catalina Mountains, 6000 ft. Connecticut: CT1, Hartford Co., West Suffield. Florida: FL1, Alachua Co., 8 km E Gainesville (E shore of Newnans Lake); FL2, Franklin Co., 7 km S Sopchoppy (E bank of Ochlocknee River); FL3, Highlands Co., Archbold Biological Station; FL4, Lake Co., Ocala National Forest, 8 km N Altoona; FL5, Walton Co., 7.5 km W Bruce; FL6, Walton Co., 6 km E Bruce, W bank of Choctawatchee River. Georgia: GA1, Greene Co., Oconee National Forest, 6 km N Greensboro; GA2, Montgomery Co., 14 km N Mount Vernon; GA3, Rabun Co., Black Rock Mountain State Park; GA4, Toombs Co., Vidalia. Indiana: IN1, Bartholomew Co., 20 km SW Columbus; IN2, Orange Co., Hoosier National Forest, 2 km SE Paoli, Pioneer Mothers Memorial Forest; IN3, Perry Co., Hoosier National Forest, 3.5 km SE Oriole. Kentucky: KY1, Laurel Co., Daniel Boone National Forest, Bald Rock Recreation Area; KY2, Rowan Co., Morehead, Daniel Boone National Forest, Rodburn Hollow Recreation Area. Maryland: MD1, Calvert Co., 8 km S Prince Frederick; MD2, Washington Co., near Hagerstown. Missouri: MO1, Callaway Co., Graham Cave State Park; MO2, St. Louis Co., Eureka. North Carolina: NC1, Davie Co., 4 km N Mocksville; NC2, Forsyth Co., Rural Hall; NC3, Haywood Co., Pisgah National Forest, 8 km SE Waterville; NC4, Iredell Co., Duke Power State Park; NC5, Montgomery Co., Uwharrie National Forest, 8 km SW Troy; NC6, Moore Co., 3 km N Southern Pines; NC7, Moore Co., Whispering Pines; NC8, Stanly Co., Morrow Mountain State Park; NC9, Swain Co., 8 km E Bryson City; NC10, Swain Co., Great Smoky Mountains National Park, Newfound Gap, 5048 ft; NC11, Stone Mountain State Park. New Jersey: NJ1, Burlington Co., Bass River State Forest; NJ2, Burlington Co., Wharton State

Table 2 (*concluded*).

Forest, 1 km E Atsion; NJ3, Mercer Co., Princeton; NJ4, Middlesex Co., New Brunswick; NJ5, Warren Co., Worthington Forest. New York: NY1, Saratoga Co., Saratoga Spa State Park. Ohio: OH1, Lawrence Co., Wayne National Forest, 4 km E Pedro, Lake Vesuvius Recreation Area, Sand Hill; OH2, Morgan Co., Wayne National Forest, 20 km W Stockport, Wildcat Hollow. Pennsylvania: PA1, Butler Co., Morraine State Park; PA2, Cumberland Co., Dickinson; PA3, Perry Co., 8.7 km S Millersburg, Notch Road, 4.5 km W Hwy 11-15; PA4, Schuylkill Co., 4 km W Pine Grove; PA5, Sullivan Co., Wyoming State Forest. South Carolina: SC1, Beaufort Co., Hunting Island State Park; SC2, Berkeley Co., Francis Marion National Forest; SC3, Chesterfield Co., Carolina Sandhills National Wildlife Refuge; SC4, Colleton Co., Colleton State Park. Tennessee: TN1, Cumberland Co., 5 km W Crab Orchard; TN2, Knox Co., Knoxville. Texas: TX1, Travis Co., vicinity of Austin (the single colony from here had been reared at the University of Texas for at least 3 years). Virginia: VA1, Roanoke Co., Roanoke. West Virginia: WV1, Braxton Co., Servia; WV2, Preston Co., 13 km W Bruceton Mills, Coopers Rock State Forest.

Note: Karyotypes were verified except those identified as follows: *, karyotyped but not verified; †, not karyotyped.

dependent. On average, the most heavily sculpted form is *fulva*; for example, the rugosity on the forecoxa of the *fulva* worker illustrated in Fig. 3*a* (also Fig. 3*b*, but the contrast is poor) is heavier than for any illustrated workers of the other forms (Figs. 3*d*–3*h*, 4*a*–4*h*). However, the relief of this sculpture decreases with decreasing body size. Three *fulva* workers (Figs. 1*a*–1*c*) demonstrate the reduction in cephalic sculpture as head width declines (the change in head shape is also apparent); note that the workers shown in Figs. 1*b* and 1*c* are from the same colony. In fact, the sides of the alitrunk of the *fulva* minim (Fig. 3*c*) are smooth and shiny, with much less sculpture than any of the other alitrunks in Figs. 3 and 4; the *fulva* minim also lacks rugose sculpture on the forecoxa. Also note the difference in cephalic sculpture between large (Fig. 2*d*) and small (Fig. 2*e*) workers of N21b from Arizona. The outstanding feature of the head of the *punctaticeps* paratype (Fig. 2*f*) is not that it is so lightly sculpted and posteriorly pointed (MacKay 1989), but that such a large individual possesses these characteristics. For sculpture to be useful, it is necessary to account for its high size dependency, but there is likely little information here that is not more readily obtained from standard quantitative characters.

Creighton (1950) considered colour to be highly unreliable for identification, but it must be examined like any other character, and its limits of variability assessed. In the *fulva*–*rudis*–*texana* complex, several forms are relatively constant in colour, and thus the dark brown (often appearing black to the naked eye) N17 and N18 (*picea*) are readily distinguished from the much lighter yellowish brown to reddish brown N20 (*carolinensis*), N19, and N22b, and still separated fairly readily from the medium brown N16 and the more variable (light brown to a fairly dark medium brown) N22a (*rudis*). However, N21a (*miamiana*) has a wide range of colour variation correlated with geographic distribution (medium reddish brown in southern Florida but light brown to dark brown elsewhere), and N21b (*texana*) has an extremely wide colour range (yellow to dark brown). Further details on colour are provided in the key.

Other qualitative differences in the workers tend to be slight or exasperatingly inconsistent. For example, Creighton (1950) characterized *fulva* by the strong projection of the anterior edge of the mesonotum above the pronotum and by the long propodeal spines being directed more strongly upwards rather than backwards. Many *fulva* workers possess these characters (both are illustrated fairly well in Fig. 3*b*), but they are not sufficiently consistent and nonoverlapping with other forms to be diagnostic on their own. Certainly the *fulva* minim (Fig. 3*c*) lacks the strong mesonotal projection, but it is also not much stronger in the largest *fulva* worker illustrated (Fig. 3*a*) than in the N16 and N17 workers illustrated (Figs. 3*d*, 3*e*). As Crozier (1977) noted, there is often a rather strong mesonotal projection in N18 workers. Similarly, there is a zone of overlap in the angle of the propodeal spines, particularly with the longer spined forms N17, N18, and N21a.

The most distinctive taxonomic character in the complex is found on the *fulva* queen (Carroll 1975). The mesopleurae (anepisternum and katepisternum) are heavily rugose in *fulva* queens (Fig. 5*a*) but largely smooth and shining in *rudis*–*texana* subcomplex queens (e.g., Fig. 5*b*). Carroll (1975) noted that males of *fulva* had thick propodeal spines, while "*carolinensis*" (N20 and N21a, I believe) rarely had propodeal spines. The propodeal armature of the males of other forms is somewhat variable; most have only small nubs but N16 usually has thick propodeal spines somewhat like those of *fulva*. I have also briefly examined representative male genitalia for each form. There may be potentially useful characters here, primarily in the digitus and cuspis, but if so the differences are not glaringly obvious. As with other morphological studies on the reproductives, this is an avenue for further investigation.

Multivariate analysis of the regular data set

Table 3 presents summary statistics, including each form's sample size and mean, minimum, and maximum observations and coefficient of variation (CV) for each variable. Table 4 gives pooled within-class and between-classes correlation matrices.

Logarithmic (base e) transformation made little difference in the overall numbers misclassified in the linear discriminant analyses; the results were only marginally better when the data set was transformed (51 were misclassified; see Table 5) than when it was not (55 were misclassified). However, a comparison of the individual correlation matrices for three forms with good sample sizes (Umphrey 1992) showed that most two-variable correlations were higher when the data set was logarithmically transformed (out of 66 correlation coefficients, higher values for transformed data numbered 42 for *fulva*, 63 for N20, and 62 for N22a). For this reason, the multivariate results discussed are based on the analysis of log-transformed data. For simplicity, the data

Table 3. Summary univariate statistics for each form in the *fulva–rudis–texana* complex of *Aphaenogaster*, regular data set.

Variable	Statistic	*fulva* (n = 38)	N16 (n = 15)	N17 (n = 33)	N18 (n= 39)	N19 (n = 5)	N20 (n = 32)	N21a (n = 13)	N21b (n = 13)	N22a (n = 67)	N22b (n = 9)
HW	Mean	0.99	1.00	0.95	0.97	0.82	0.87	0.94	0.91	0.90	1.03
	Min.	0.81	0.80	0.82	0.80	0.71	0.64	0.85	0.75	0.70	0.84
	Max.	1.16	1.12	1.15	1.11	0.97	1.05	1.03	1.06	1.07	1.18
	CV	8.31	8.84	7.58	7.57	12.55	11.53	5.81	10.68	9.02	11.86
HL	Mean	1.20	1.21	1.17	1.18	1.12	1.15	1.22	1.27	1.16	1.28
	Min.	1.04	1.02	1.06	1.03	1.03	0.86	1.12	1.08	0.97	1.10
	Max.	1.36	1.34	1.36	1.32	1.24	1.32	1.32	1.44	1.32	1.41
	CV	6.13	7.01	6.02	5.82	7.25	8.61	4.42	8.76	6.05	8.38
SL	Mean	1.30	1.28	1.35	1.34	1.34	1.32	1.41	1.49	1.32	1.38
	Min.	1.12	1.14	1.23	1.20	1.26	1.06	1.32	1.31	1.16	1.21
	Max.	1.46	1.38	1.51	1.44	1.47	1.46	1.52	1.64	1.46	1.48
	CV	5.71	5.33	4.24	4.02	6.27	6.72	4.67	6.82	5.07	6.99
EL	Mean	0.24	0.25	0.24	0.24	0.24	0.24	0.28	0.24	0.24	0.25
	Min.	0.20	0.20	0.21	0.20	0.20	0.18	0.24	0.20	0.18	0.22
	Max.	0.28	0.27	0.30	0.27	0.28	0.30	0.30	0.28	0.27	0.28
	CV	7.52	8.25	7.63	7.52	12.57	10.30	5.72	9.26	7.89	8.81
WLA	Mean	1.61	1.59	1.61	1.62	1.53	1.55	1.65	1.71	1.55	1.70
	Min.	1.40	1.34	1.42	1.42	1.42	1.20	1.52	1.46	1.28	1.46
	Max.	1.78	1.72	1.88	1.80	1.72	1.76	1.76	1.92	1.74	1.84
	CV	6.02	6.61	6.18	5.71	7.89	8.29	4.02	8.24	6.10	8.38
MH	Mean	1.01	1.03	1.02	1.05	0.96	0.98	1.05	1.09	0.99	1.10
	Min.	0.84	0.84	0.86	0.86	0.86	0.70	0.96	0.88	0.76	0.92
	Max.	1.20	1.12	1.24	1.20	1.10	1.14	1.14	1.24	1.14	1.22
	CV	7.55	7.92	7.67	7.29	9.97	9.75	4.98	11.24	7.56	10.40
SPL	Mean	0.34	0.27	0.31	0.31	0.26	0.27	0.34	0.25	0.26	0.29
	Min.	0.22	0.19	0.27	0.24	0.22	0.18	0.28	0.20	0.19	0.22
	Max.	0.42	0.33	0.40	0.40	0.31	0.33	0.40	0.30	0.32	0.36
	CV	12.38	12.09	10.20	9.48	13.70	13.14	11.31	11.45	11.89	15.67
ISPL	Mean	0.21	0.14	0.17	0.17	0.12	0.14	0.20	0.09	0.14	0.15
	Min.	0.11	0.09	0.14	0.13	0.09	0.06	0.16	0.05	0.07	0.10
	Max.	0.28	0.18	0.22	0.24	0.16	0.18	0.25	0.12	0.19	0.20
	CV	17.14	16.20	12.72	11.94	22.82	18.86	13.10	21.57	18.67	23.03
SPD	Mean	0.34	0.30	0.31	0.32	0.26	0.28	0.33	0.30	0.28	0.28
	Min.	0.26	0.22	0.24	0.25	0.22	0.20	0.30	0.25	0.22	0.23
	Max.	0.43	0.35	0.40	0.39	0.30	0.33	0.36	0.38	0.36	0.32
	CV	13.19	9.57	12.60	10.76	14.39	11.15	6.05	13.32	9.94	11.06
FL	Mean	1.67	1.56	1.60	1.60	1.59	1.59	1.71	1.85	1.57	1.65
	Min.	1.43	1.30	1.41	1.39	1.44	1.24	1.54	1.55	1.30	1.44
	Max.	1.90	1.68	1.83	1.77	1.76	1.79	1.88	2.10	1.74	1.76
	CV	6.14	6.38	5.83	5.40	7.56	8.19	5.93	8.99	5.76	7.37
FW	Mean	0.23	0.26	0.25	0.25	0.24	0.24	0.25	0.27	0.25	0.27
	Min.	0.20	0.22	0.22	0.22	0.22	0.18	0.23	0.22	0.21	0.24
	Max.	0.26	0.29	0.28	0.28	0.26	0.28	0.27	0.32	0.28	0.29
	CV	6.88	7.21	5.96	6.07	6.43	8.79	5.59	11.12	6.30	6.14
TL	Mean	1.15	1.06	1.08	1.08	1.10	1.08	1.18	1.29	1.07	1.14
	Min.	0.98	0.86	0.95	0.94	0.99	0.84	1.06	1.06	0.87	0.96
	Max.	1.28	1.14	1.26	1.21	1.24	1.22	1.32	1.48	1.18	1.20
	CV	6.36	6.82	6.04	5.19	8.61	8.29	6.09	9.64	5.75	7.75

Note: Mean, sample mean; min., minimum observation in sample; max., maximum observation in sample; CV, coefficient of variation (%). All measurements are given in millimetres.

Fig. 1. Frontal views of heads of workers of the *fulva–rudis–texana* complex of *Aphaenogaster*. Data are given as the form code or species name, followed by the colony number, locality code or locality, head width (HW) in millimetres, and Weber's length of alitrunk (WLA) in millimetres. (*a*) *Aphaenogaster fulva*, A168, NJ3; HW = 1.11, WLA = 1.78. (*b* and *c*) *A. fulva*, A385, IN2; HW = 1.00, WLA = 1.62, and HW = 0.63, WLA = 1.21, respectively. (*d*) N16, A188, ON5; HW = 0.99, WLA = 1.64. (*e*) N17, A204, ON5; HW = 1.04, WLA = 1.77. (*f*) N18 (*picea*), A203, ON1; HW = 1.05, WLA = 1.81. (*g*) N16 (?), A380; Mo.: Franklin Co., Shaw Arboretum; HW = 0.76, WLA = 1.30. (*h*) N19, A372, MO1; HW = 0.87, WLA = 1.65. (*i*) N20 (*carolinensis*), A179, NJ1; HW = 1.00, WLA = 1.73.

Fig. 2. Frontal views of heads of workers of the *fulva–rudis–texana* complex of *Aphaenogaster*. Data are given as the form code or species name, followed by the colony number, locality code or locality, head width (HW) in millimetres, and Weber's length of alitrunk (WLA) in millimetres. (*a*) N21a (*miamiana*), A096, FL3; HW = 0.87, WLA = 1.64. (*b*) N21a (*miamiana*), A242, FL1; HW = 1.00, WLA = 1.77. (*c*) N21b (?), GJU0327; Brownsville, Tex.; HW = 0.97, WLA = 1.83. (*d*) N21b (*texana*), GJU1046; Chiricahua National Monument, Herb Martyr Forest Camp, Cochise Co., Ariz., 6150′; HW = 1.00, WLA = 1.87. (*e*) N21b (*texana*), A437, AZ1; HW = 0.71, WLA = 1.40. (*f*) *A. punctaticeps*, unique paratype; Dona Ana Co., N.Mex., 45 km NE Las Cruces, Jornado Experimental Range; HW = 1.04, WLA = 1.90. (*g*) N22a (*rudis*), A085, AL1; HW = 0.83, WLA = 1.49. (*h*) N22b, A222, ON6; HW = 0.95, WLA = 1.62. (*i*) N17 × N22b hybrid, A221, ON6; HW = 1.04, WLA = 1.77.

Fig. 3. Lateral views of alitrunks of workers of the *fulva–rudis–texana* complex of *Aphaenogaster*. Individuals illustrated are those shown in Figs. 1*a*–1*f*, 1*h*, and 1*i*, respectively. Data are given as the form code or species name, followed by the colony number, locality code or locality, and Weber's length of alitrunk (WLA) in millimetres. (*a*) *Aphaenogaster fulva*, A168, NJ3; WLA = 1.78. (*b* and *c*) *A. fulva*, WLA = 1.62 and 1.21, respectively. (*d*) N16, A188, ON5; WLA = 1.64. (*e*) N17, A204, ON5; WLA = 1.77. (*f*) N18 (*picea*), A203, ON1; WLA = 1.81. (*g*) N19, A372, MO1; WLA = 1.65. (*h*) N20 (*carolinensis*), A179, NJ1; WLA = 1.73.

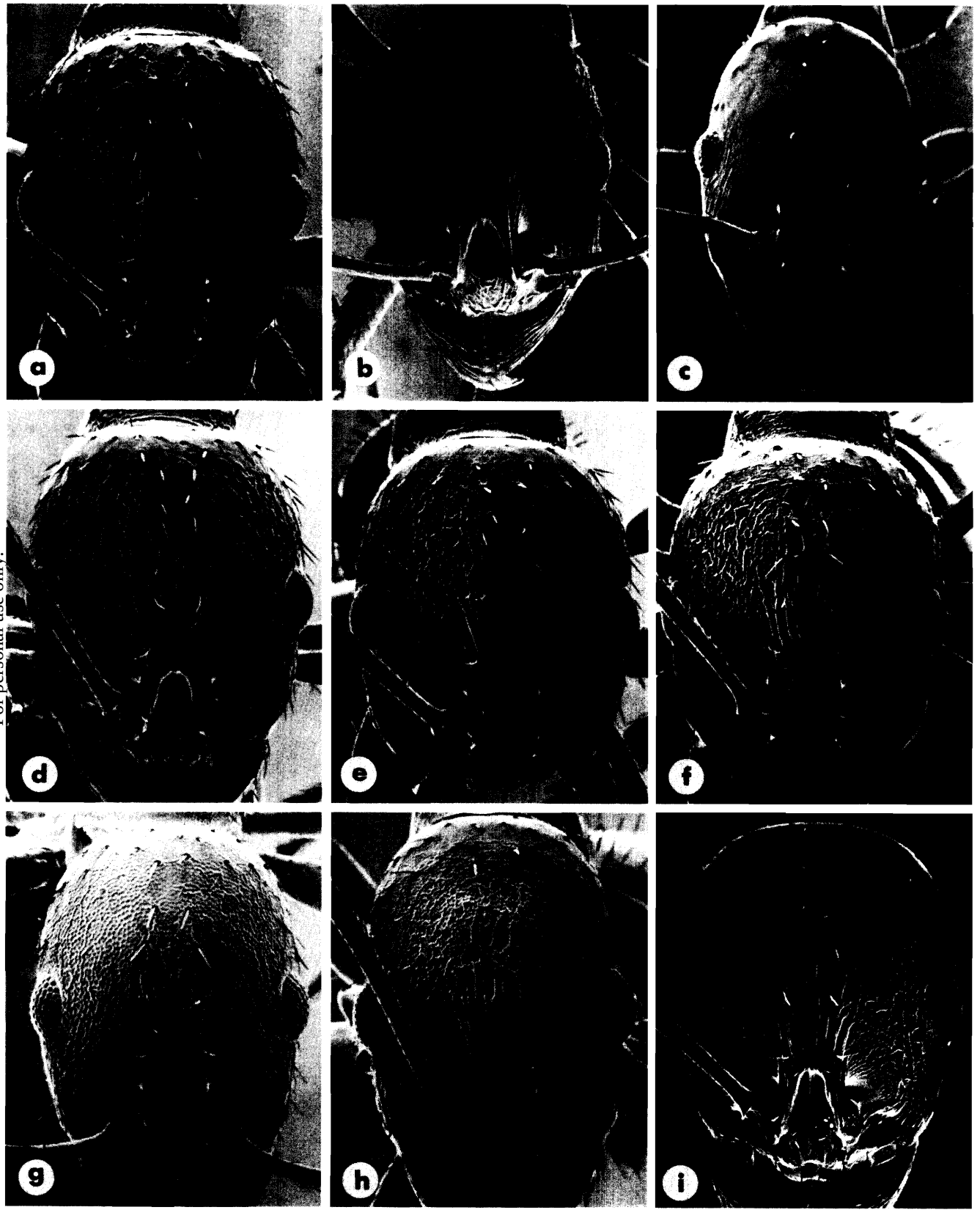
Table 4. Pooled within-class correlation matrix (top) and between-classes correlation matrix (bottom) for the regular data set, log transformed.

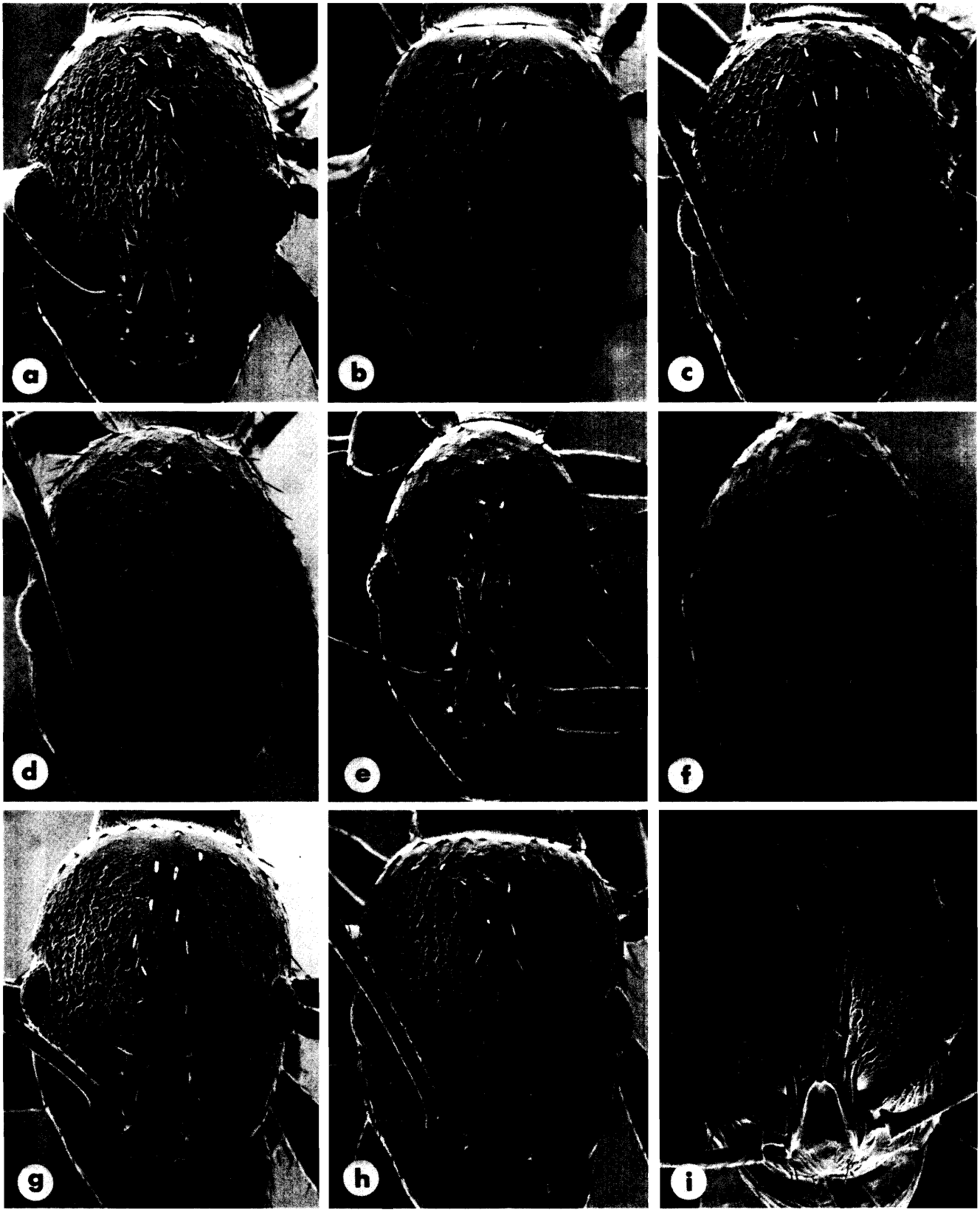
	HW	HL	SL	EL	WLA	MH	SPL	ISPL	SPD	FL	FW	TL
HW	1	0.970	0.903	0.915	0.962	0.949	0.856	0.824	0.751	0.920	0.898	0.902
HL		1	0.928	0.905	0.971	0.954	0.862	0.830	0.754	0.948	0.913	0.934
SL			1	0.850	0.923	0.891	0.798	0.770	0.679	0.939	0.849	0.930
EL				1	0.910	0.908	0.793	0.772	0.708	0.881	0.856	0.870
WLA					1	0.964	0.861	0.829	0.762	0.951	0.905	0.933
MH						1	0.845	0.806	0.773	0.930	0.883	0.909
SPL							1	0.935	0.702	0.840	0.811	0.834
ISPL								1	0.660	0.811	0.781	0.798
SPD									1	0.728	0.732	0.697
FL										1	0.886	0.969
FW											1	0.867
TL												1
HW	1	0.650	−0.102	0.223	0.638	0.660	0.703	0.612	0.716	0.211	0.086	0.134
HL		1	0.532	0.510	0.918	0.871	0.302	0.049	0.432	0.762	0.453	0.710
SL			1	0.413	0.676	0.630	−0.070	−0.360	0.016	0.750	0.606	0.708
EL				1	0.468	0.460	0.348	0.271	0.305	0.441	0.187	0.433
WLA					1	0.971	0.431	0.124	0.531	0.781	0.487	0.701
MH						1	0.372	0.092	0.465	0.632	0.604	0.539
SPL							1	0.934	0.908	0.276	−0.412	0.232
ISPL								1	0.787	−0.033	−0.569	−0.056
SPD									1	0.447	−0.340	0.417
FL										1	0.166	0.991
FW											1	0.092
TL												1

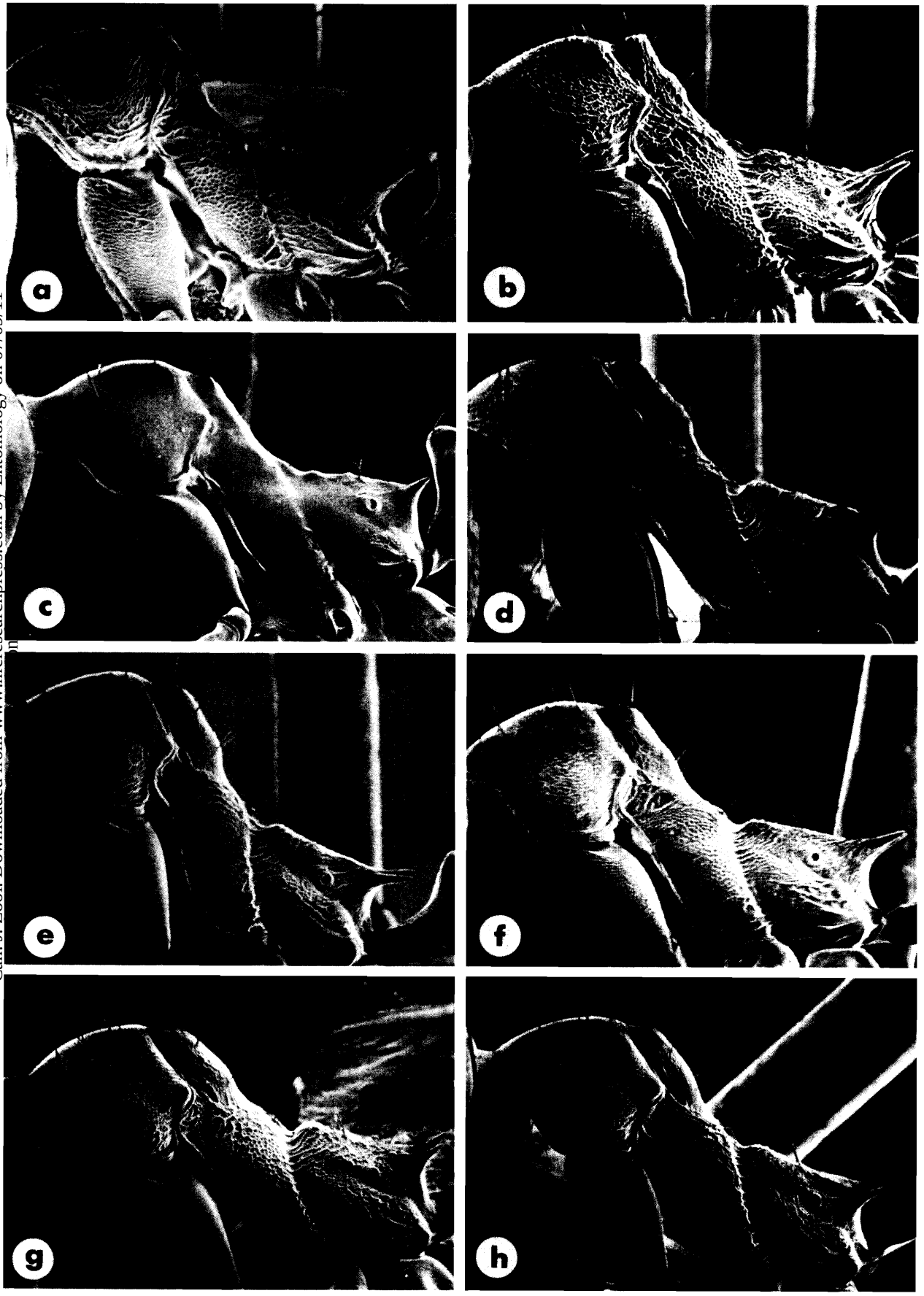
Note: The total sample consisted of 264 observations among 10 classes.

set was not transformed for the two-variable scatterplots (Figs. 13–48).

The most serious misclassifications in the linear discriminant analysis on the regular data set (Table 5) occurred between the two dark brown forms, N17 and N18. Since N19 has not been found within the range of N20 sampled, the misclassification between these two forms may not be a practical taxonomic problem. The perfect classification of N21a is misleading, as discussed later, but discrimination of *fulva* workers was excellent; posterior probabilities of belonging to *fulva* were 0.0000 for all non-*fulva* workers and 1.0000 for all *fulva* workers except one (0.9981 for A303). Note







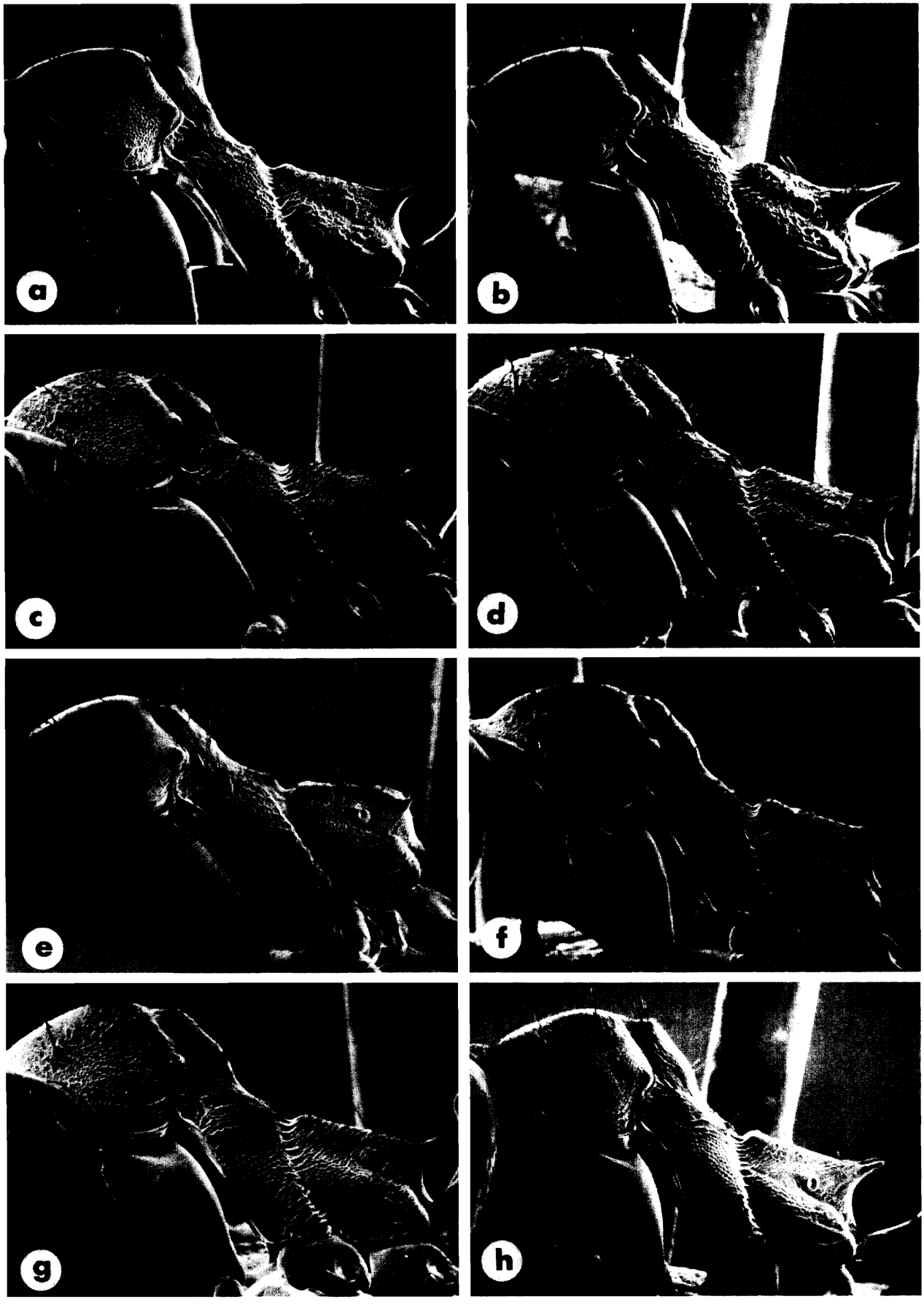


Fig. 4. Lateral views of alitrunks of workers of the *fulva–rudis–texana* complex of *Aphaenogaster*. Individuals illustrated are those shown in Figs. 2*a*–2*d* and 2*f*–2*i*, respectively. Data are given as the form code or species name, followed by the colony number, locality code or locality, and Weber's length of alitrunk (WLA) in millimetres. (*a*) N21a (*miamiana*), A096, FL3; WLA = 1.64. (*b*) N21a (*miamiana*), A242, FL1; WLA = 1.77. (*c*) N21b (?), GJU0327; Brownsville, Texas; WLA = 1.83. (*d*) N21b (*texana*), GJU1046; Chiricahua National Monument, Herb Martyr Forest Camp, Cochise Co., Arizona, 6150 ft; WLA = 1.90. (*e*) *A. punctaticeps*, unique paratype; Dona Ana Co., New Mexico, 45 km NE Las Cruces, Jornado Experimental Range; WLA = 1.90. (*f*) N22a (*rudis*), A085, AL1; WLA = 1.49. (*g*) N22b, A222, ON6; WLA = 1.62. (*h*) N17 × N22b hybrid, A221, ON6; WLA = 1.77.

Fig. 5. Lateral view of the mesopleurae (anepisternum and katepisternum) and surrounding area of the alitrunk on queens of the *fulva–rudis–texana* complex of *Aphaenogaster*. Data are given as the form code or species name, followed by the colony number and locality code. (*a*) *fulva*, A385, IN2. (*b*) N16, A409, ON5.

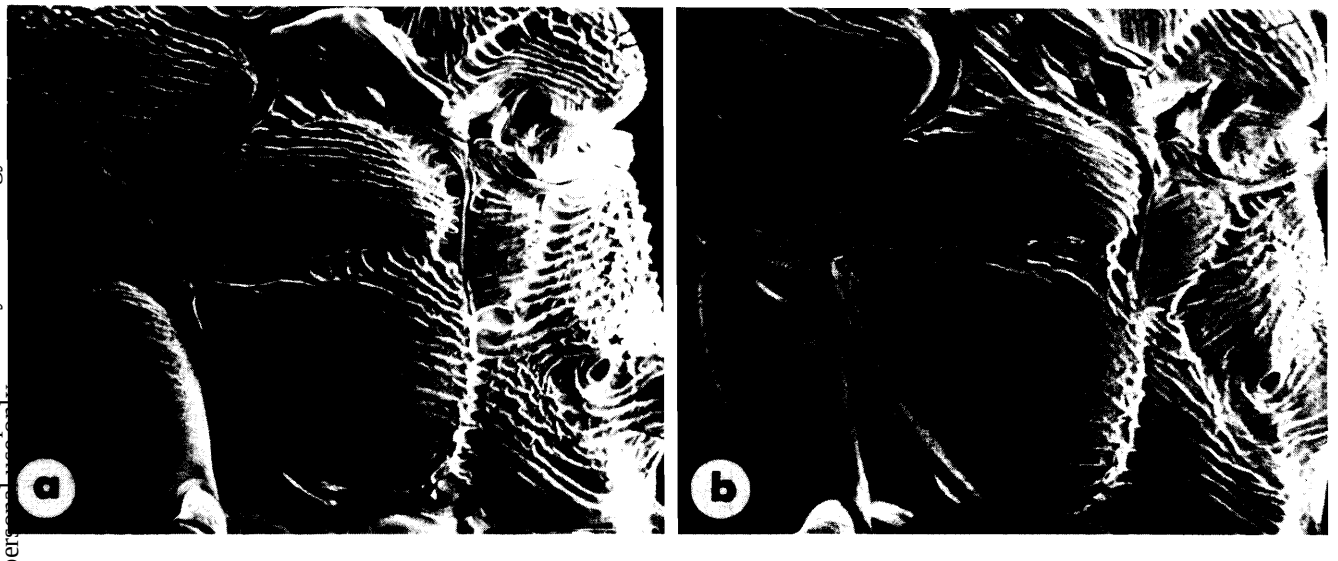


Table 5. Classification results (using jackknife) for linear discriminant analysis, regular data set, log transformed, for 10 forms of the *fulva–rudis–texana* complex of *Aphaenogaster*.

	Number of individuals classified into form:									
From	*fulva*	N16	N17	N18	N19	N20	N21a	N21b	N22a	N22b
fulva	38 (100.0)	0	0	0	0	0	0	0	0	0
N16	0	13 (86.7)	0	0	0	0	0	0	1 (6.7)	1 (6.7)
N17	0	0	27 (81.8)	6 (18.2)	0	0	0	0	0	0
N18	0	0	14 (35.9)	24 (61.5)	0	0	0	0	1 (2.6)	0
N19	0	0	0	0	4 (80.0)	1 (20.0)	0	0	0	0
N20	0	0	0	0	6 (18.8)	22 (68.8)	0	0	4 (12.5)	0
N21a	0	0	0	0	0	0	13 (100.0)	0	0	0
N21b	0	0	0	0	1 (7.7)	0	0	12 (92.3)	0	0
N22a	0	0	6 (9.0)	1 (1.5)	1 (1.5)	5 (7.5)	0	0	51 (76.1)	2 (3.0)
N22b	0	0	0	0	0	0	0	0	1 (11.1)	8 (88.9)

Note: Values in parentheses are percentages.

Fig. 6. Plot of the first two canonical variates from the canonical variate analysis of the log-transformed regular data set with all forms included. The small graphs in *b*–*e* contain subsets of the forms in order to give a breakdown of the graph in *a* with all forms plotted.

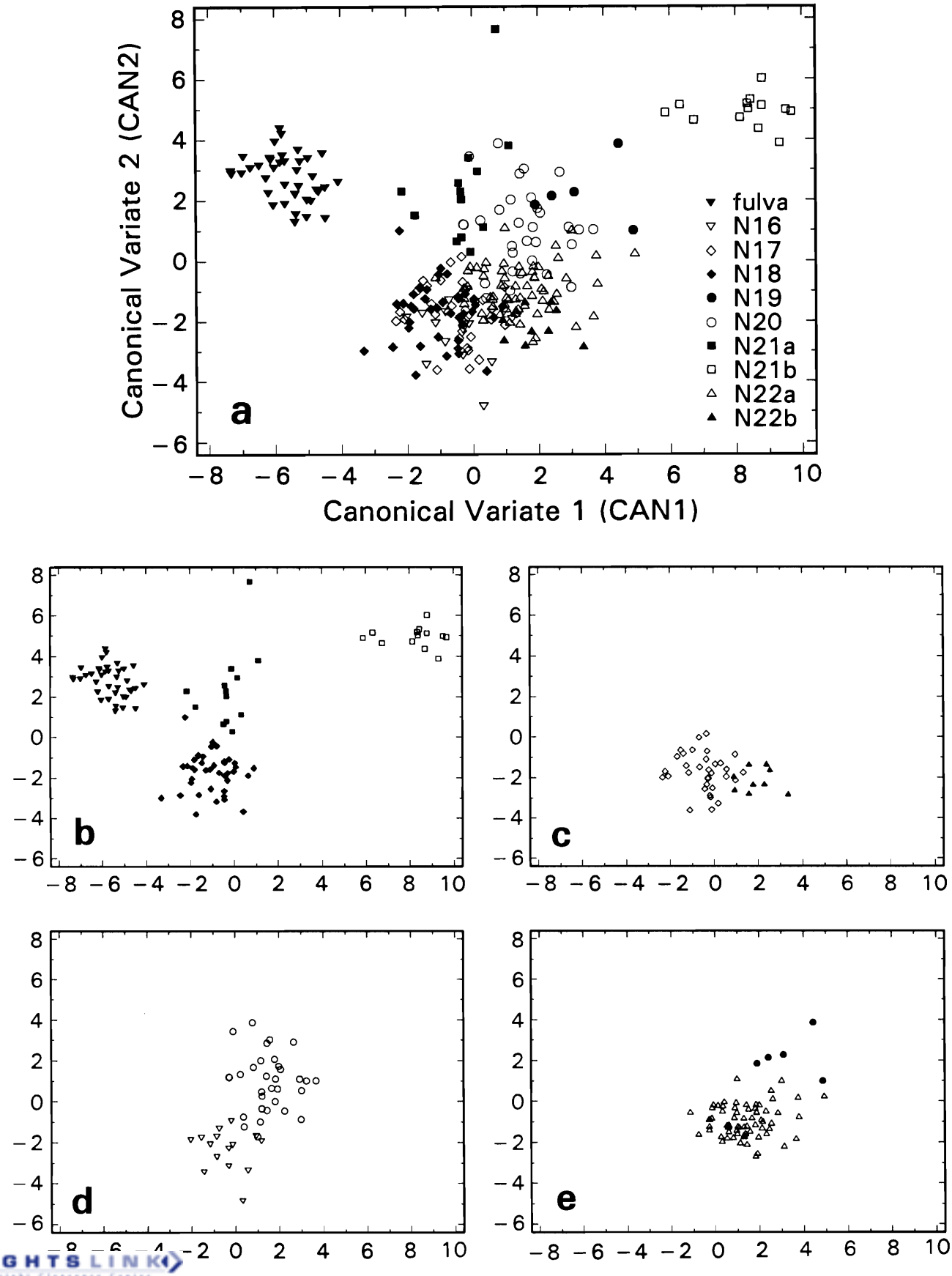


Table 6. Selected results of canonical variate analysis for the *fulva–rudis–texana* complex in *Aphaenogaster*; log-transformed regular data set with all 10 forms included and with *fulva* and N21b excluded.

	(*a*) With all 10 forms included			(*b*) With *fulva* and N21b excluded		
	CAN1	CAN2	CAN3	CAN1	CAN2	CAN3
Raw canonical coefficients (and constant)						
Constant	2.790	−25.283	35.867	14.971	4.863	25.942
HW	−25.827	−14.837	−22.345	−30.602	13.739	10.923
HL	11.837	17.653	−19.310	5.085	−44.465	19.725
SL	18.756	−13.990	32.691	16.515	18.630	−30.990
EL	2.672	4.737	17.671	18.643	−7.193	14.844
MH	8.988	−13.413	4.622	−6.843	−0.492	−14.338
WLA	6.179	−1.254	−2.167	−1.654	11.043	−12.223
SPL	2.383	3.603	1.877	2.326	6.756	−1.977
ISPL	−6.981	−3.118	4.810	2.266	3.186	2.995
SPD	−3.758	3.229	1.479	1.698	4.663	4.076
FL	−5.285	19.071	−8.422	2.727	−5.624	−3.074
FW	16.412	−17.229	−0.263	−8.584	−4.183	−11.836
TL	−5.452	22.985	−8.810	6.639	−3.055	9.552
Standardized canonical coefficients						
HW	−2.69	−1.54	−2.32	−3.19	1.43	1.34
HL	0.86	1.29	−1.41	0.36	−3.16	1.40
SL	1.18	−0.88	2.06	0.93	1.05	−1.74
EL	0.24	0.42	1.58	1.71	−0.66	1.36
MH	0.78	−1.18	0.41	−0.59	−0.04	−1.24
WLA	0.44	−0.09	−0.15	−0.12	0.77	−0.86
SPL	0.38	0.58	0.30	0.34	1.00	−0.30
ISPL	−1.96	−0.88	1.35	0.51	0.71	0.67
SPD	−0.51	0.44	0.20	0.21	0.58	0.51
FL	−0.40	1.45	−0.64	0.18	−0.38	−0.21
FW	1.34	−1.41	−0.02	−0.61	−0.30	−0.85
TL	−0.44	1.87	−0.72	0.46	−0.21	0.66
% var	51.2	23.9	13.8	49.6	34.4	8.5
Class means of canonical variables						
fulva	−5.64	2.77	−1.11			
N16	−0.43	−2.34	−3.04	−3.45	−1.61	1.98
N17	−0.47	−1.73	1.07	−0.27	2.05	−0.38
N18	−1.02	−1.75	0.17	−1.12	2.11	0.00
N19	3.35	2.21	1.96	3.30	−1.59	−1.81
N20	1.52	0.94	1.32	1.98	−1.18	−0.02
N21a	−0.29	2.40	4.19	4.95	0.92	1.79
N21b	8.23	4.94	−2.49			
N22a	1.35	−0.93	−0.11	−0.28	−1.07	−0.42
N22b	1.92	−2.12	−2.25	−2.37	−2.30	−0.33

Note: % var is the percentage of among-groups variation explained by the particular canonical variable.

also that N22a, which I believe is the true *rudis*, gets confused with several other forms.

Canonical variate analysis on the regular data set

A plot of the first two canonical variates (Fig. 6) from the canonical variate analysis on the regular data set (with all forms included) illustrates the morphological distinctness of *fulva*. Good separation of *fulva* and N21b from the other eight forms requires only the first canonical variate (CAN1), [illegible]onical variate (CAN2) reduces the chances of misclassifying *fulva* and N21b. The other eight forms are plotted in a rather amorphous mass between CAN1 values of −4 and 5, but there is a pattern here as well. For example, all N21a workers have CAN2 > 0, while all N16 and N22b workers have CAN2 < 0. The N21a observation at CAN2 = 7.65 is a properly measured individual from colony A096 collected at FL3 (Archbold Biological Station), and often appears as an outlier; workers in this colony have unusually narrow heads and long thin femora.

Of the nine canonical variates generated in the analysis of

Fig. 7. Canonical variate 2 on canonical variate 1 for the log-transformed regular data set with *fulva* and N21b deleted from the analysis.

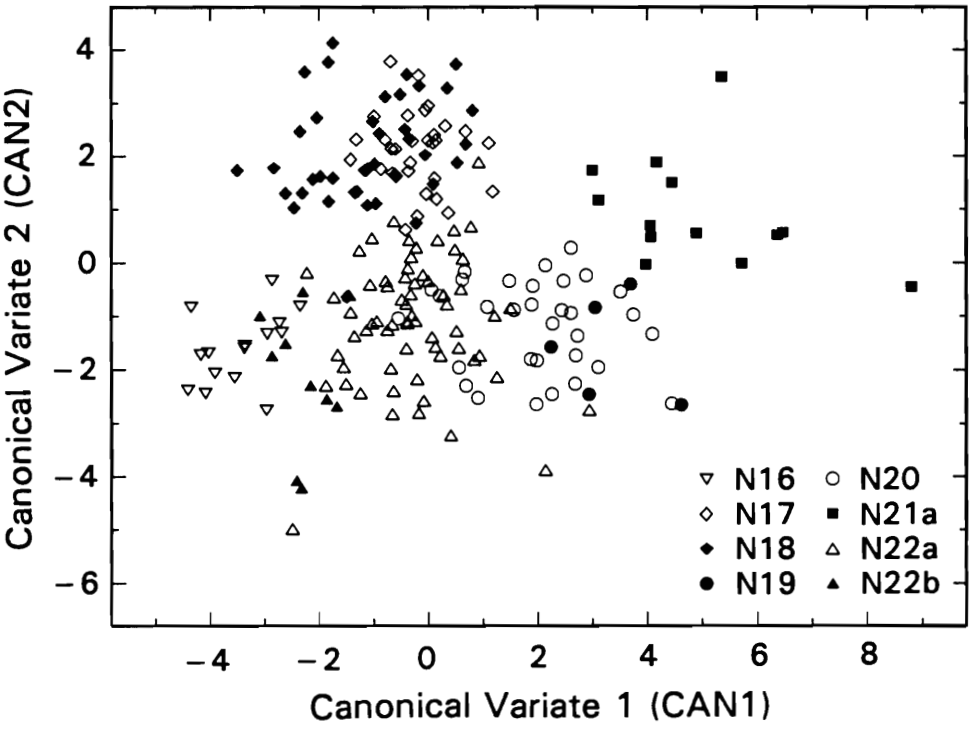
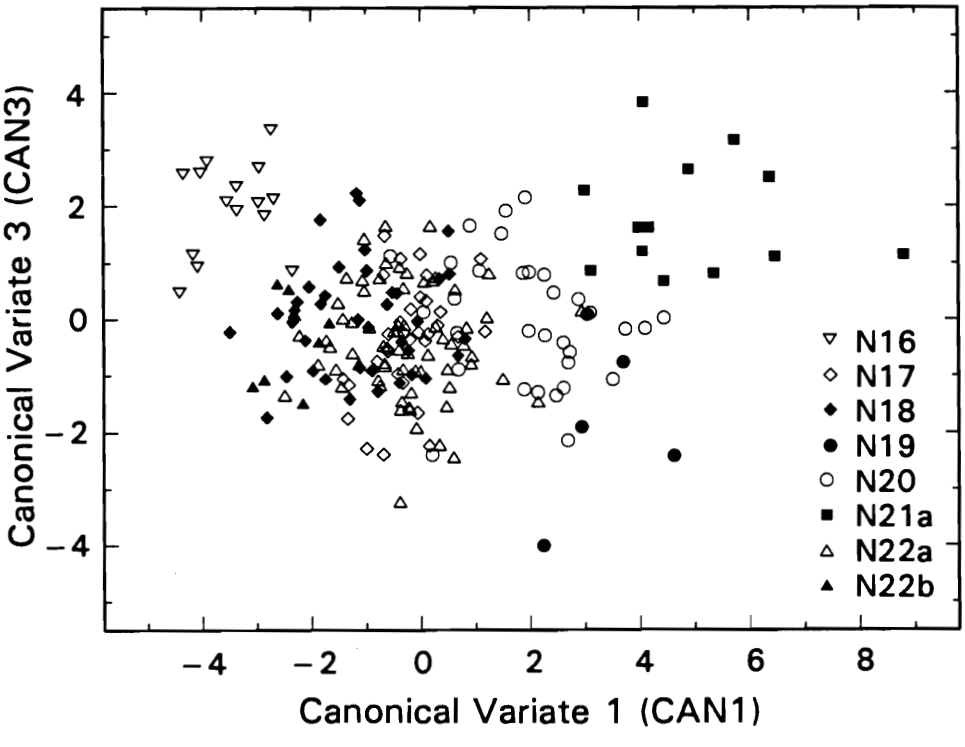


Fig. 8. Canonical variate 3 on canonical variate 1 for the log-transformed regular data set with *fulva* and N21b deleted from the analysis.



all forms, the first four account for 96.9% of the among-groups variation; CAN1 alone accounts for 51.2% and CAN2 accounts for 23.9% (Table 6*a*). Some groups are separated further on other canonical variates, but it is often difficult to see this: the presence of *fulva* and N21b partly obscures differentiation among the other forms.

To enhance visual separation, another canonical variate analysis was made on a reduced data set obtained by removing *fulva* and N21b. In the plot of CAN2 versus CAN1 (Fig. 7), five or six fairly distinct zones can be discerned. Four of the zones are composed mainly of N21a (the most distinct zone), N20 overlapping N19, N22a, and N17 over-

Table 7. Results of canonical variate analysis on three subsets of forms from the *fulva–rudis–texana* complex of *Aphaenogaster*.

	(*a*) N16, N22, N22b		(*b*) N17, N18, N22a		(*c*) N20, N21a, N22a	
	CAN1	CAN2	CAN1	CAN2	CAN1	CAN2
Raw canonical coefficients (and constant)						
Constant	−16.115	34.980	0.609	30.316	19.583	24.782
HW	−24.688	22.313	22.560	−16.840	−15.170	18.080
HL	−9.438	−29.431	−40.468	−4.272	−5.850	−10.830
SL	31.772	−11.571	13.175	29.327	3.058	17.727
EL	−2.566	8.304	−7.216	13.613	21.165	−6.549
MH	2.841	−7.826	3.708	−15.155	−8.732	−10.171
WLA	−27.740	−27.947	11.037	−7.971	1.489	−8.514
SPL	3.012	−4.434	10.483	10.672	6.135	−8.323
ISPL	0.430	2.965	−0.651	−0.406	1.214	8.621
SPD	3.270	9.489	3.164	−4.092	5.110	5.746
FL	36.335	12.211	−7.381	10.538	1.496	−27.383
FW	−5.741	−3.801	−5.795	8.854	−16.805	4.819
TL	7.421	10.463	−12.076	−13.165	7.490	12.094
Standardized canonical coefficients						
HW	−2.65	2.39	2.03	−1.51	−1.52	1.82
HL	−0.68	−2.13	−2.45	−0.26	−0.42	−0.78
SL	1.79	−0.65	0.62	1.38	0.19	1.07
EL	−0.22	0.71	−0.57	1.07	2.12	−0.66
MH	0.25	−0.68	0.30	−1.21	−0.75	−0.87
WLA	−1.95	−1.96	0.70	−0.51	0.10	−0.60
SPL	0.39	−0.58	1.49	1.51	0.92	−1.24
ISPL	0.09	0.60	−0.13	−0.08	0.29	2.06
SPD	0.34	0.98	0.39	−0.50	0.60	0.67
FL	2.29	0.77	−0.43	0.61	0.11	−1.98
FW	−0.40	−0.27	−0.36	0.55	−1.23	0.35
TL	0.48	0.67	−0.69	−0.76	0.56	0.91
% var	83.0	17.0	90.2	9.8	91.3	8.7
Class means of canonical variables						
N16	−3.77	0.94				
N17			1.39	0.85		
N18			1.73	−0.65		
N20					0.87	−0.90
N21a					4.62	0.82
N22a	1.11	0.12	−1.69	−0.04	−1.31	0.27
N22b	−2.01	−2.45				

Note: Data are log transformed. % var is the percentage of among-groups variation explained by the particular canonical variable.

lapping N18. In addition, N16 is distinct from all except N22b, which more or less fits between N16 and N22a. In the plot of CAN3 versus CAN1 (Fig. 8) there is better separation of N16 and N22b, but one N16 point is still quite close to some N22b points. Of the seven canonical variates generated, the first three account for 92.5% of the among-groups variation (Table 6*b*).

Separate canonical variate analyses were performed on three selected triplets of forms; each included N22a, as this form is the most widespread and abundant, and most likely to be confused with another form. Analyzing three forms can generate only two canonical variates, so plotting CAN2 against CAN1 illustrates all among-groups variation (see Table 7 for a breakdown). Although there will be some misclassifications using only measurement data, N16, N22a, and N22b are reasonably distinct (Fig. 9). The plot for N17, N18, and N22a (Fig. 10) shows substantial overlap between N17 and N18, and both overlap with N22a to some extent. The plot for N20, N21a, and N22a (Fig. 11) shows fairly considerable overlap between N20 and N22a.

Mahalonobis' generalized distances

The most commonly used measure of distance between two multivariate means is Mahalanobis' generalized distance, which corrects for the effect of correlation among variables (Lachenbruch 1975). The population parameter is denoted by δ^2, which is estimated by the sample statistic D^2. Following Baker et al. (1972), I have given D, the square root of

Fig. 9. Plot of the two canonical variates using only the data on N16, N22a, and N22b.

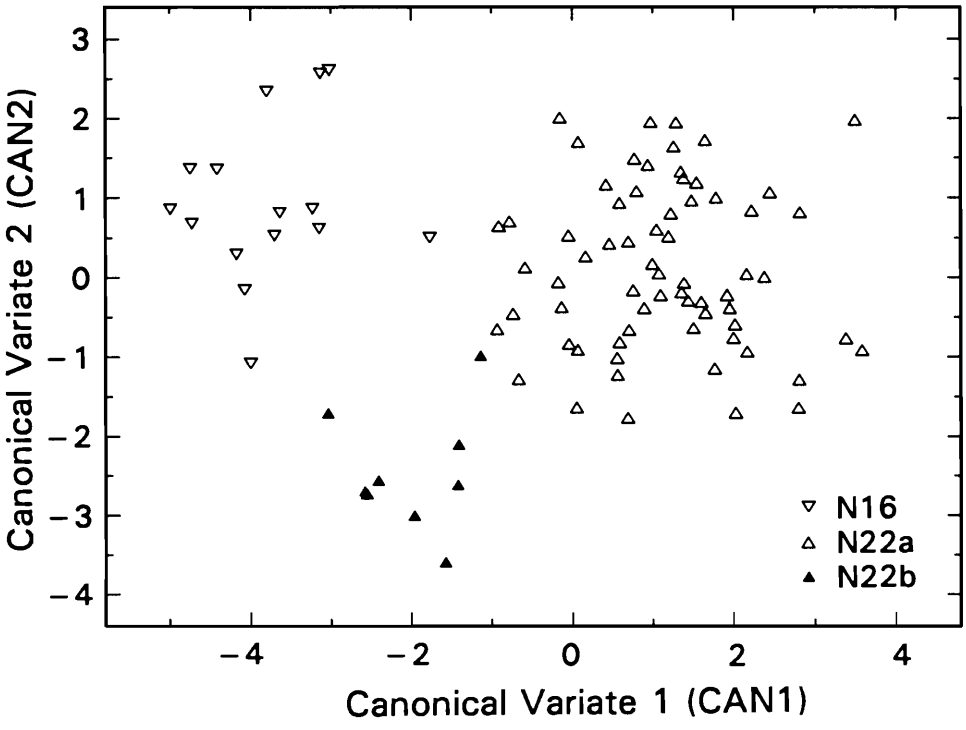
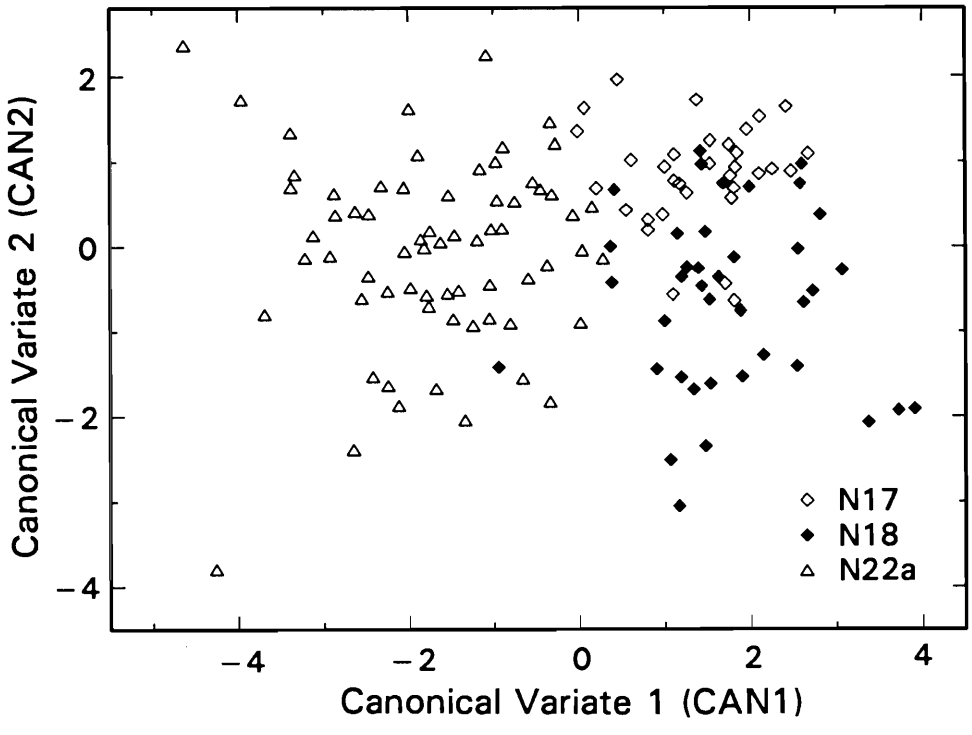


Fig. 10. Plot of the two canonical variates using only the data on N17, N18, and N22a.



Mahalanobis' generalized distance, rather than the more usually reported D^2 (Table 8). These values can be more readily interpreted in terms of the standard normal distribution. Let Φ denote the cumulative standard normal distribution function. Then if equal prior probabilities are used, the probability of misclassification between two groups from approximately multivariate normal populations with multivariate means separated by the distance δ^2 is approximately $\Phi(-\delta/2)$ (Lachenbruch 1975). To obtain a probability of misclassification that is less than 0.01, δ must be at least 4.66; less than 0.001 requires δ be at least 6.19. Only one observation was misclassified between forms separated by $D \geq 4.66$ (see Tables 3 and 6): an observation of N21b was classified as N19, even though their D value was 7.61.

Fig. 11. Plot of the two canonical variates using only the data on N20, N21a, and N22a.

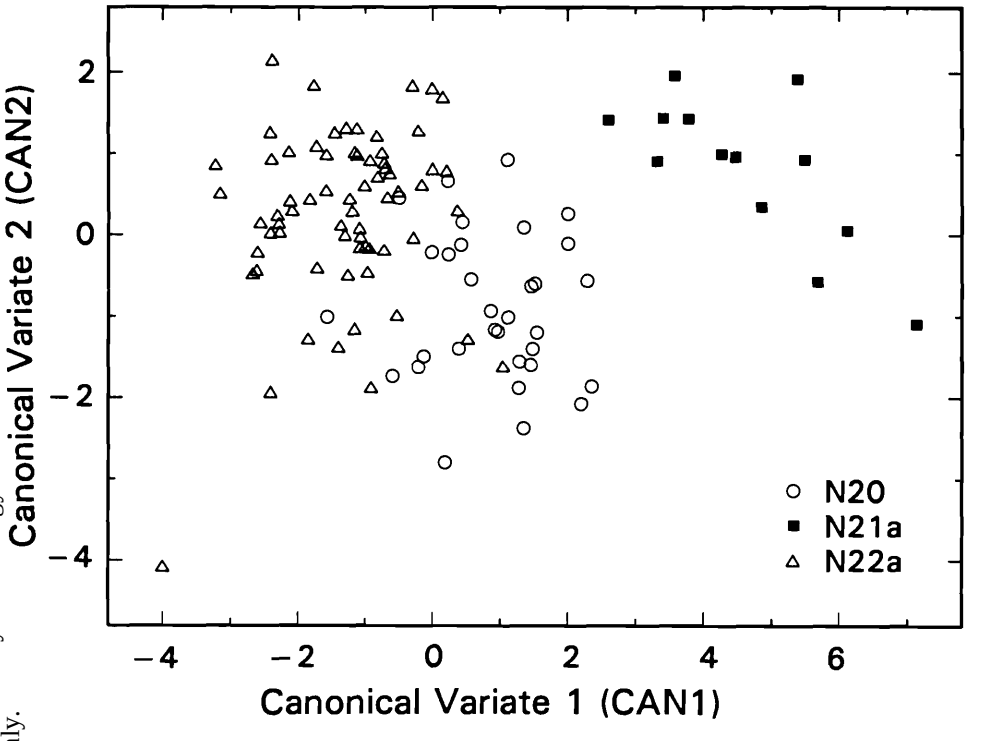


Table 8. Square roots of Mahalanobis' generalized distances (D^2) between forms of the *fulva–rudis–texana* complex of *Aphaenogaster* (regular data set, log transformed).

	fulva	N16	N17	N18	N19	N20	N21a	N21b	N22a	N22b
fulva	0	7.86	7.35	6.81	9.59	7.88	7.74	14.25	8.02	9.40
N16	7.86	0	5.39	4.88	8.16	5.99	8.72	11.98	4.17	3.64
N17	7.35	5.39	0	1.50	5.95	4.26	5.92	11.55	3.28	5.31
N18	6.81	4.88	1.50	0	6.53	4.76	6.57	11.74	3.57	5.27
N19	9.59	8.16	5.95	6.53	0	2.74	4.93	7.61	4.52	6.76
N20	7.88	5.99	4.26	4.76	2.74	0	4.04	9.25	2.63	5.16
N21a	7.74	8.72	5.92	6.57	4.93	4.04	0	11.59	5.93	8.45
N21b	14.25	11.98	11.55	11.74	7.61	9.25	11.59	0	9.77	10.24
N22a	8.02	4.17	3.28	3.57	4.52	2.63	5.93	9.77	0	3.63
N22b	9.40	3.64	5.31	5.27	6.76	5.16	8.45	10.24	3.63	0

Small sample sizes of 5 for N19 and 13 for N21b may account for this (see Lachenbruch 1975).

The results for forms separated by D values of less than 4.66 are consistent with the probabilities obtained from substituting D for δ into the cumulative standard normal distribution function expression given above. The results are complicated by some forms being close to more than one other form; for example, the D value is less than 4.66 between N22a and six other forms. Nonetheless we can consider the results of those correctly classified or misclassified into the other form when comparing two forms in order to obtain an empirical misclassification rate (undoubtedly biased to some degree). For example, D = 2.63 between N20 and N22a gives an estimated probability of misclassification of 0.093. For N20, 22 specimens were classified correctly and 4 were misclassified as N22a; for N22a, 51 specimens were classified correctly and 5 were misclassified as N20. The empirical misclassification rate is thus 9/85 = 0.106. Similarly, between N17 and N18 (the two closest forms) we obtain an empirical misclassification rate of 20/71 = 0.282 compared with a theoretical rate of 0.227. In each case, the empirical and theoretical rates do not differ statistically (χ^2 goodness-of-fit test; $p > 0.10$).

Stepwise discriminant analysis

Some pairs of variables are highly correlated, both within and between classes (Table 4); these include (using an arbitrarily chosen value of 0.9) the obvious pairs of FL and TL, ISPL and SPL, and MH and WLA, as well as HL and WLA. Several other pairs of variables have high within-class (although not between-class) correlations; note the values between WLA and each of HW, SL, EL, FL, FW, and TL. Weaker within-class correlations between WLA and the three spinal variables likely reflect highly variant spinal development and (or) nonlinearity rather than a lack of size dependency or measurement error.

Table 9. Classification results (using jackknife) for linear discriminant analysis, regular data set log transformed, using reduced set of seven variables, for 10 forms of the *fulva–rudis–texana* complex of *Aphaenogaster*.

	Number of individuals classified into form:									
From	*fulva*	N16	N17	N18	N19	N20	N21a	N21b	N22a	N22b
fulva	38 (100.0)	0	0	0	0	0	0	0	0	0
N16	0	14 (93.3)	0	0	0	0	0	0	0	1 (6.7)
N17	0	0	23 (69.7)	9 (27.3)	0	0	0	0	1 (3.0)	0
N18	0	0	12 (30.8)	25 (64.1)	0	0	0	0	1 (2.6)	1 (2.6)
N19	0	0	0	0	3 (60.0)	1 (20.0)	1 (20.0)	0	0	0
N20	0	0	0	0	9 (28.1)	17 (53.1)	0	0	6 (18.8)	0
N21a	0	0	0	0	0	1 (7.7)	12 (92.3)	0	0	0
N21b	0	0	0	0	0	0	0	13 (100.0)	0	0
N22a	0	0	5 (7.5)	6 (9.0)	1 (1.5)	6 (9.0)	0	0	44 (65.7)	5 (7.5)
N22b	0	0	0	0	0	0	0	0	1 (11.1)	8 (88.9)

Note: Values in parentheses are percentages.

Stepwise discriminant analysis of all 10 forms entered the variables in the order ISPL, TL, FW, HW, EL, SL, HL, SPL, SPD, MH, FL, WLA. Each variable made a statistically significant addition to the average squared canonical correlation, although only marginally so for WLA ($p = 0.0434$ compared with $p \leq 0.0001$ for entry of each of the other variables). No variables could be removed at any stage. Entry of ISPL first indicates that it is the single best discriminator (the second to fourth best choices as the first variable entered were SPL, TL, and SPD, respectively). Emery's (1895) decision to classify members of this complex on the basis of the length of the propodeal spines is thus seen to be quite reasonable. It is also symptomatic of the difficulties in morphologically separating members of this complex: the single best variable for discrimination also has the highest relative variation, as measured by the coefficient of variation. Each form has a higher CV for ISPL than for any other variable (Table 3).

Since most of the information contained in SPL has already been accounted for by ISPL, it enters the model late (eighth). Similarly, TL (the second variable entered) explains almost all information accounted for by FL (the 11th variable entered). It makes little difference which of TL or FL is used, but TL may be more easily and precisely measured. The very late entries of MH and WLA are not surprising, as they are simply measures of body size; this information is contained in other variables. I consider WLA to be the single best variable for "body size" comparisons among species in the complex, and adequate within species. Both HW and HL are more easily and precisely measured for body size comparisons within species.

After all variables were entered, the last step attempted to remove a variable from the model. This step is no longer necessary, given that the other variables were in the model. The five most difficult variables to remove were (in order, starting with the most difficult) EL, HW, SL, FW, and ISPL, with partial R^2 values for removal of 0.472, 0.468, 0.443, 0.409, and 0.386, respectively. This indicates which variables contain the most unique information for discriminating among groups. The appearance of ISPL in this group is surprising, since it would be expected that SPL would contain most of the information contained by ISPL, but SPL has a partial R^2 value for removal of only 0.202. In contrast, it is fairly easy to remove either TL or FL, given that the other is in the model. This suggests that ISPL really is a more discriminating variable than SPL.

Although the entry of a variable into the model may have a "statistically significant" effect in further separating the group means, it may be unnecessary for purposes of discrimination. The first seven variables into the model gave an average squared correlation of 0.339, while adding the other five variables raised this value to only 0.393. Hence, a discriminant analysis was run using only the first seven variables to be entered into the model (ISPL, TL, FW, HW, EL, SL, HL). For this analysis, 67 individuals were misclassified (Table 9) compared with 51 with all variables used (Table 5). The discrimination of *fulva*, N16, and N21b was still excellent, but there is substantially more confusion among the other forms. At the current stage of model development I recommend measuring all 12 variables to help separate the phenetically close forms.

Interpretation of standardized canonical coefficients

The absolute value of a standardized canonical coefficient is proportional to its contribution to the canonical variate. In the canonical variate analysis of all 10 forms, the largest

coefficients in absolute value for the first canonical variate (CAN1) are those for HW and ISPL, and they have the same sign (Table 6*a*). This sets up a strong contrast between the broad-headed, long-spined *fulva* and the narrow-headed, short-spined N21b; other forms fall in between. The coefficient for SL is of the opposite sign, as the scapes of *fulva* are relatively shorter than those of N21b. Although a femur of *fulva* is relatively slightly thinner than one of N21b, the primary contribution of FW is probably to separate *fulva* from most of the other forms, which generally have a stockier femur. The second canonical variate (CAN2) is more difficult to interpret, but forms with relatively wide heads also tend to have wide femora (*fulva* is an exception), broad heads, and short legs. Note that EL makes little contribution until the third canonical variate (CAN3); also, CAN3 is placing a strong contrast between head size (HW and HL both negative) and SL (positive).

When *fulva* and N21b are excluded from the canonical variate analysis, HW is still the most important contributor to CAN1, but now EL is second in importance, although of opposite sign (Table 6*b*). This contrasts the small-eyed, broad-headed forms against the larger-eyed, more narrow-headed forms. The largest contributor to CAN2 is HL. The standardized coefficients for the three spinal measurements do not look very large, but CAN2 is highly correlated (correlation coefficient ≥ 0.82) with each of ISPL, SPL, and SPD in the among-groups canonical structure. This suggests that CAN2 has captured the discriminatory power of the spinal characters using other, less variable characters. As in the analysis with all forms, CAN3 is contrasting (at least in part) scape length against head size.

The canonical variate analysis of N16, N22a, and N22b essentially separates N16 and N22b from N22a using CAN1, and then separates N16 and N22b on CAN2 (Fig. 9). The standardized canonical coefficients (Table 7*a*) show that CAN1 contrasts head and body size (particularly HW and WLA) against scape and femur lengths; relative to N22a, N16 and N22b have short scapes and femora for their body size and head width. Although CAN2 is based heavily on HW, HL, and WLA, its interpretation is not clear. In a stepwise discriminant analysis, the order of variable entry was HW, SL, FL, HL, SPD, WLA. No other variables could be entered at the 15% level of significance (the highest partial R^2 to enter was 0.043 for SPL) and none of the six variables entered could be removed at the 1% level of significance (the lowest partial R^2 to remove was 0.133). By far the most difficult variable to remove from the model was SL, with a partial R^2 to remove of 0.647 (the next largest was 0.223 for HW). The very large size of many N22b workers likely accounts for the value of WLA.

Almost all separation in the canonical variate analysis of N17, N18, and N22a occurs on CAN1, which differentiates N22a from N17 and N18 (Fig. 10). The principal contributors to CAN1 are clearly HW, HL, and SPL (Table 7*b*). This contrasts forms with relatively short, broad heads and long spines (N17 and N18) against N22a, which has, on average, a narrower head and shorter spines for a fixed body size. Although N17 and N18 are poorly separated on CAN2, the scapes of N17 are probably, on average, longer for a fixed head width; the other high standardized coefficients (for EL, MH, and SPL) are hard to interpret.

The separation of N21a from N20 and N22a on CAN1 (Fig. 11) in the canonical analysis of these three forms is based largely on the relatively larger eyes of N21a (Table 7*c*); N21a also has relatively long narrow femora and long spines. Most of the rather poor separation of N20 and N22a is on CAN2; the tendency of N22a to have relatively longer spines and shorter femora is reflected in the values of the standardized canonical coefficients for ISPL and FL (but TL has an opposite sign to FL!). The high values of the HW coefficients for both CAN1 and CAN2 suggest that it is being used as a size contrast variable, but this is not clear. In a stepwise discriminant analysis, the order of entry of variables into the model was ISPL, FW, EL, HW, SPD, SPL, MH, FL; no other variables could be entered and none were removed at any step. The first three variables clearly contribute the most to the average squared canonical correlation: 0.394 for the first three variables versus 0.509 for all eight.

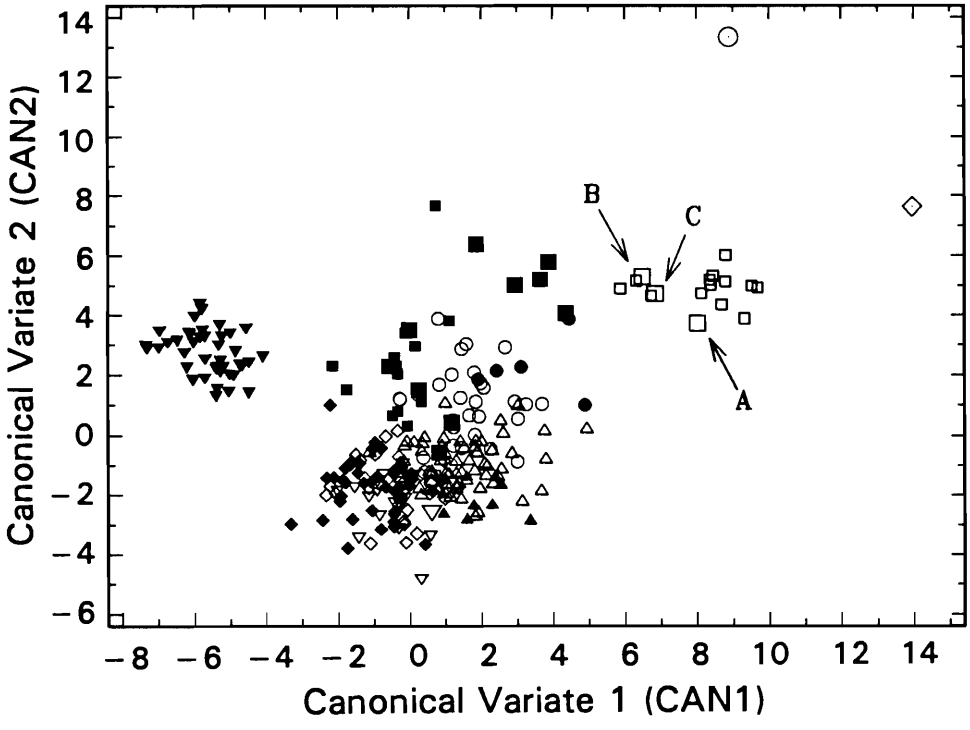
Multivariate classification of selected test data

The first two canonical variates (using the functions derived on the regular data set, all forms) were calculated on the test data set and overlaid on the regular data set plot of CAN2 on CAN1 (see Figs. 6*a*, 12; several points are obscured). Points for the paratypes of *huachucana* and *punctaticeps* (Fig. 12) are far removed from all other forms; these are incorrectly classified as N21b by the discriminant functions with posterior probability of 1.000, since N21b is the form to which they are closest. There is no doubt that *huachucana* is a distinct species, and it seems likely that *punctaticeps* must be as well; additional material of *punctaticeps* since acquired supports this view (S.P. Cover and G.J. Umphrey, in preparation). In contrast, the data point representing the paratype of *crinimera* is in very close proximity to the N21b cluster (Fig. 12, point C), consistent with the hypothesis that *crinimera* is conspecific with N21b.

Two workers identified as N21b prior to classification with the multivariate methods also are clearly in the N21b cluster in the plot of CAN2 on CAN1 (Fig. 12, points A and B). One (Arizona: Cochise Co., Miller Canyon, 5600–5800 ft (1 ft = 0.3048 m), Huachuca Mountains, coll. R.R. Snelling) was classified as N21b by means of the discriminant functions (also see Fig. 12, point A), but the other, from Brownsville, Tex., was classified as N19. This worker has CAN3 = 3.448, which is substantially outside of the observed N21b range of −4.36 to −0.13 and closer to the N19 range of 1.11–2.66 (the range for N21a is 2.17–6.27).

All 13 workers of N21a (*miamiana*) in the regular data set were properly allocated, but the performance of the discriminant functions in classifying 10 other workers that I believe are N21a was poor. Three were classified as N21a, three as N20, and four as N19. Most of the test colonies came from localities outside of the range of most colonies used in the regular data set. Three of these colonies were collected at Baton Rouge, Louisiana, in a very wet habitat, and clearly resemble other N21a colonies from northern Florida and Georgia. However, only one of these colonies was correctly identified by the discriminant functions; the other two were classified as N20. Five of the colonies (all four classified as N19 and one as N21a) came from southern Florida: one from Orlando, two from Everglades National Park, and two from Big Pine Key. These five cluster in the region between the point for colony A096 (largest CAN2 value for N21a) and the N21b points in Fig. 12. Despite

Fig. 12. Plot of the first two canonical variates from the canonical variate analysis of the log-transformed regular data set with all forms of the *fulva–rudis–texana* complex of *Aphaenogaster* included; this figure is the same as Fig. 6*a*, but an additional 30 data points have been plotted. These data points represent additional measured workers for which canonical variates were calculated using the functions generated by the canonical analysis on the log-transformed regular data set. For details see Fig. 6*a*; larger symbols represent additional individuals identified prior to the multivariate analysis. Two additional symbols represent a *huachucana* paratype (a large open diamond with a small dot inside near the right border) and the *punctaticeps* paratype (a large open circle with a dot inside near top border). "A," "B," and "C" denote specific individuals referred to in the text.



some interference from other points, it is obvious that N21a is a very heterogeneous form, the workers from colonies from southern Florida having a substantially different morphology from those farther north.

Crozier (1977) obtained a karyotype considered to be very similar to that of *fulva* on a colony from Alexander Bay, Florida, that was identified as "*rudis*-group (?'*miamiana*')." A randomly selected worker from three vouchers (kindly supplied by Ross Crozier) was classified as *fulva* with a posterior probability of 1.0000, and is within the *fulva* cluster in Fig. 12 (at −5.29, 2.16), even though this worker was smaller (HW = 0.78 mm) than any in the regular data set (minimum HW = 0.81 mm). Although *fulva* and N21a appear quite distinct in the canonical variate plots (Figs. 6, 12), they are quite easy to misidentify where they are sympatric. Both have relatively long propodeal spines and relatively long thin femora and nest in wood in mesic habitats (see Carroll 1975). The SL–HW relationship (Fig. 23) is likely the easiest way to separate workers of these two forms.

James Trager (personal communication) found an *Aphaenogaster* sp. nesting in prairie habitats in Missouri (Fig. 1*g*). A karyotype could not be obtained from a colony he provided (A380; Missouri: Franklin Co., Shaw Arboretum), but a worker included in the test data set was classified as N16 with a posterior probability of 0.9929. This represents a substantial extension to the known range of N16 and is consistent with the hypothesis that N16 was originally a prairie (or at least midwestern) form.

Two-variable scatterplots

Since two-variable scatterplots present the data in a more conventional and perhaps more easily understood manner, the set of 36 scattergrams is presented in Figs. 13–48; the first 29 use variables measured directly ("conventional" two-variable plots), while the last 7 involve indices. Any conventional two-variable plot for all 10 forms shows substantial overlap between forms (e.g., Figs. 13, 21, 31, 32, 36, 38), but breaking down these plots gives diagnostic characters for separating some pairs of forms. In other cases, the plots may not provide diagnostic characters but may provide further evidence on the reliability of an identification. They also show where some of the problems occur with existing taxonomic keys and help to illustrate the nature of morphological differentiation in this complex.

In all but three plots, estimated linear regression lines have been superimposed on the scatter plots for the range of the data. These lines (*i*) illustrate the nature of interspecific variability in the two-variable relationships, (*ii*) help tie together the points for a particular form, and (*iii*) serve as a base for visually estimating intraspecific variability. It is usually clear in the plots involving small subsets of the 10 forms as to which form each line belongs to; in most cases

Figs. 13–18. Two-variable scattergrams with fitted regression lines, except that Fig. 15 gives regression lines only for data of Fig. 13.

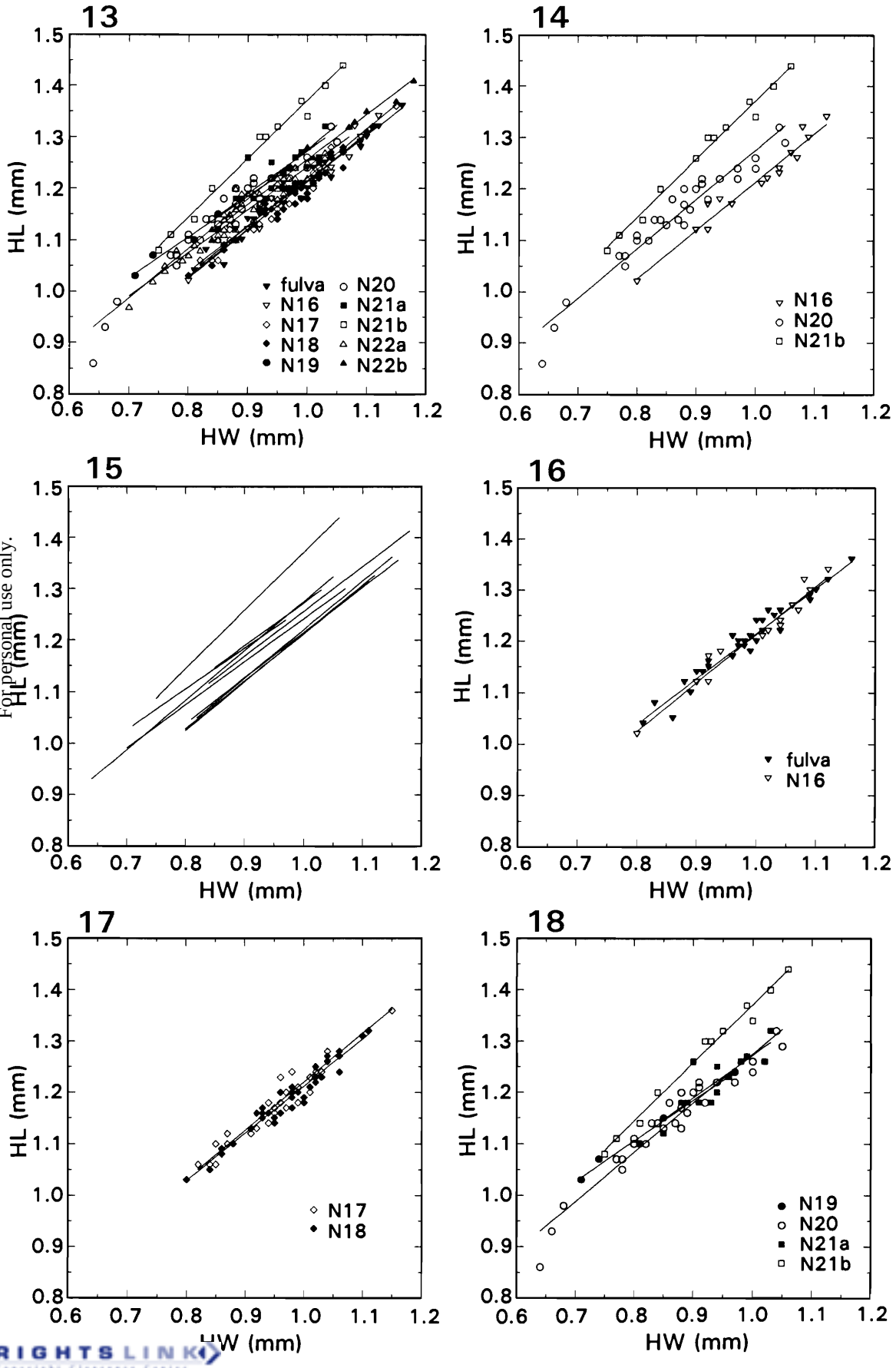


Table 10. Parameter estimates and individual standard errors for five two-variable linear regression relationships for each of the 10 forms in the *fulva–rudis–texana* complex of *Aphaenogaster*.

Y	*X*	Form	b_0	b_1	SE(b_0)	SE(b_1)
HL	HW	*fulva*	0.334	0.881	0.031	0.031
		N16	0.272	0.941	0.056	0.055
		N17	0.274	0.946	0.044	0.046
		N18	0.294	0.918	0.032	0.033
		N19	0.476	0.786	0.044	0.054
		N20	0.317	0.958	0.038	0.044
		N21a	0.437	0.835	0.150	0.160
		N21b	0.236	1.135	0.037	0.040
		N22a	0.408	0.833	0.025	0.028
		N22b	0.382	0.874	0.040	0.039
SL	HW	*fulva*	0.470	0.843	0.056	0.056
		N16	0.542	0.736	0.064	0.064
		N17	0.648	0.742	0.050	0.052
		N18	0.722	0.638	0.058	0.060
		N19	0.729	0.749	0.158	0.193
		N20	0.642	0.783	0.066	0.076
		N21a	0.624	0.841	0.246	0.262
		N21b	0.575	1.009	0.076	0.083
		N22a	0.668	0.724	0.044	0.049
		N22b	0.576	0.776	0.051	0.050
ISPL	WLA	*fulva*	−0.323	0.333	0.047	0.030
		N16	−0.120	0.162	0.062	0.039
		N17	−0.116	0.180	0.037	0.023
		N18	−0.086	0.158	0.041	0.025
		N19	−0.222	0.224	0.035	0.023
		N20	−0.148	0.184	0.024	0.015
		N21a	−0.380	0.353	0.093	0.056
		N21b	−0.091	0.105	0.044	0.026
		N22a	−0.192	0.211	0.032	0.020
		N22b	−0.253	0.238	0.037	0.022
FW	FL	*fulva*	−0.006	0.141	0.018	0.011
		N16	−0.006	0.170	0.034	0.022
		N17	−0.032	0.138	0.024	0.015
		N18	0.010	0.151	0.024	0.015
		N19	0.040	0.123	0.024	0.015
		N20	0.020	0.138	0.025	0.015
		N21a	0.096	0.091	0.054	0.032
		N21b	−0.057	0.178	0.021	0.012
		N22a	0.008	0.154	0.016	0.010
		N22b	0.055	0.130	0.026	0.016
EL	WLA	*fulva*	−0.016	0.160	0.026	0.016
		N16	−0.048	0.184	0.025	0.015
		N17	−0.035	0.172	0.020	0.012
		N18	−0.033	0.168	0.026	0.016
		N19	−0.126	0.237	0.059	0.039
		N20	−0.037	0.182	0.021	0.013
		N21a	−0.049	0.198	0.067	0.041
		N21b	−0.017	0.152	0.025	0.015
		N22a	−0.042	0.179	0.016	0.010
		N22b	0.007	0.146	0.038	0.023

Note: Estimators of the regression parameters are b_0, the *Y* intercept, and b_1, the slope. Each standard error is based on data from that form only.

this can be determined by checking the identity of the points at the ends of the regression lines. In other cases, it is essentially irrelevant to know this.

Table 10 gives estimates of regression parameters and associated standard errors for five important relationships: HL on HW, SL on HW, ISPL on WLA, FW on FL, and EL on WLA. The standard errors use an estimate of the variance about the regression line based only on the data for each form rather than a pooled estimate from all 10 forms. Estimates could also have been based on pooling the mean square error (MSE) from the ANOVA tables for testing the significance of the regressions. For those favouring such an approach, and as a quick way of making statistical comparisons, the following values for the standard deviation of *Y* about the regression line ($s_{Y.X}$, the square root of the pooled MSE) are presented: HL on HW, 0.01874; SL on HW, 0.03124; ISPL on WLA, 0.01434; FW on FL, 0.008114; EL on WLA, 0.008417.

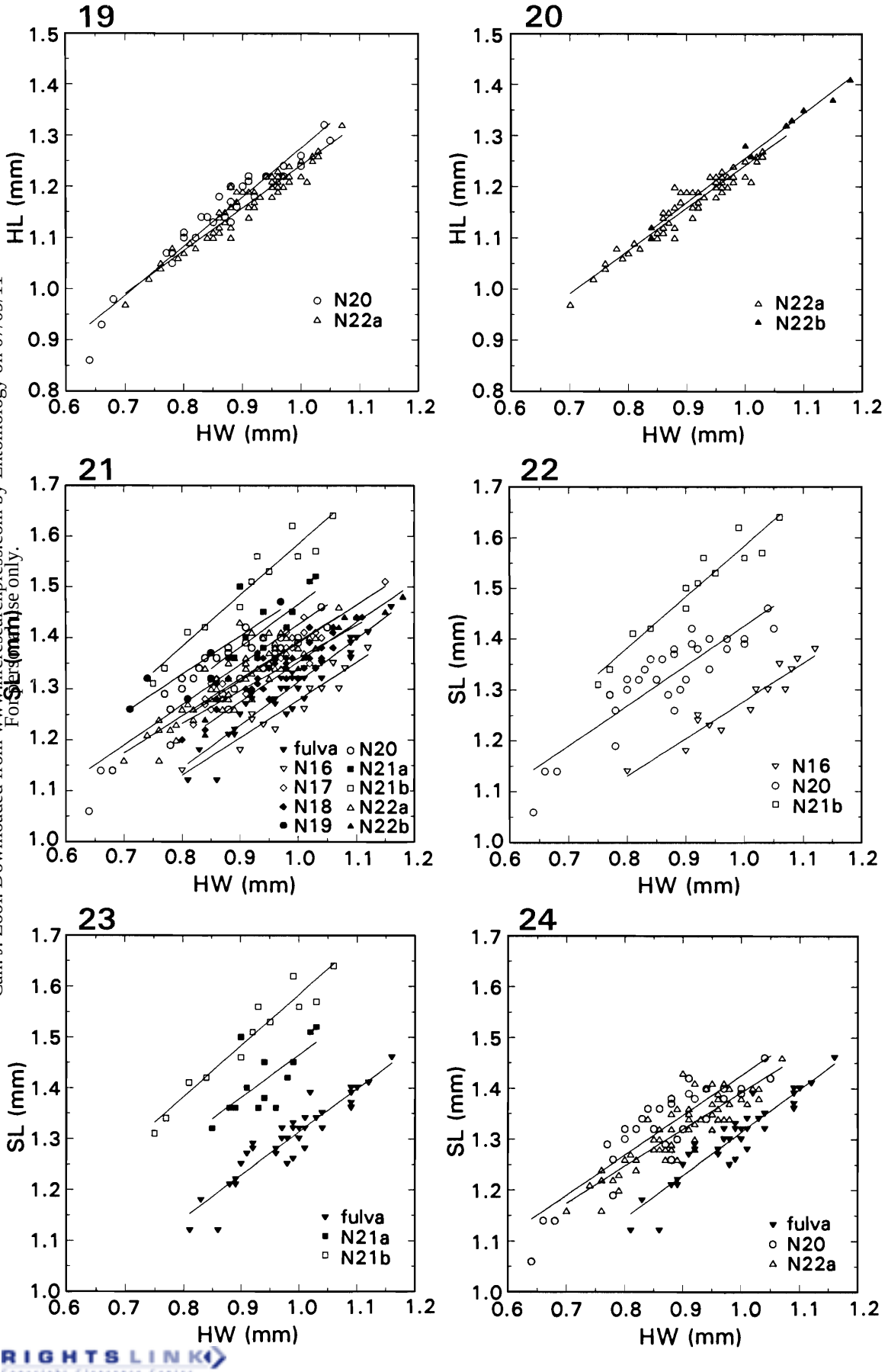
In a plot of HL versus HW (Fig. 13), only the N21b data points are reasonably distinct, while those of the other nine forms appear as a poorly differentiated mass. However, each species has a rather tight zone of variability about its fitted regression line, so that it can be said that the head of N21b is usually narrower than that of N20 (or any other form), while the head of N20 is narrower than that of N16 (Fig. 14), for a fixed value of HL.

If only the regression lines of Fig. 13 are plotted (Fig. 15), three distinct groups can be discerned: (1) the most narrow-headed group consists of N21b only; (2) five forms (N19, N20, N21a, N22a, N22b) compose a group with moderately narrow heads, although there appears to be some variability; (3) four forms (*fulva*, N16, N17, N18) compose a group with the broadest heads. The interspecific variability for head shape for the four broad-headed forms is negligible compared with the intraspecific variability for each form (Figs. 16, 17; if these diagrams are superimposed, the scatterplots overlap completely). A similar phenomenon is seen among N19, N20, and N21a (Fig. 18). However, comparing N20 and N22a (Fig. 19), it would appear that, on average, HL increases at a slower rate as HW increases (or HW increases at a faster rate as HL increases) for N22a, although there is considerable overlap. The regressions of N22a and N22b are very similar (Fig. 20), but the tendency of mature colonies of N22b to have very large workers is obvious.

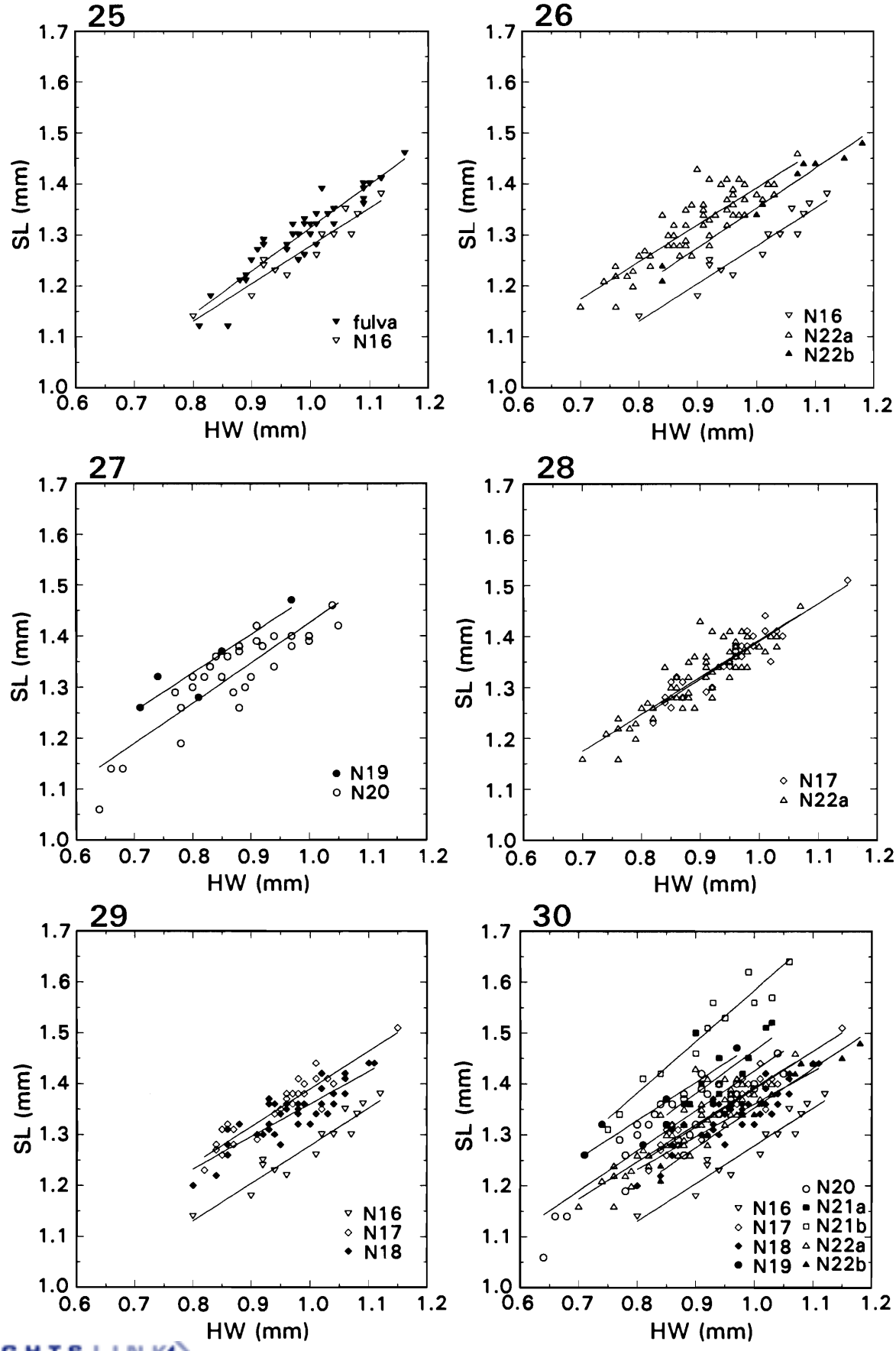
Couplet 19 of Creighton's (1950) key separated *texana* (N21b) and *carolinensis* (N20) from *miamiana*, *rudis*, and *picea* (N21a, N22a, and N18, respectively, in my usage, but likely containing other forms in Creighton's usage) on the basis of head shape. Creighton considered the latter group to have "head of the largest workers (mandibles excluded) not more than one-sixth longer than broad; head of the smaller workers approximately one-fifth longer than broad," while *texana* and *carolinensis* had "head of the worker, regardless of size, approximately one-third longer than broad." This couplet cannot do what Creighton intended; N21b does have a relatively narrower head than the other forms, but N20's HL–HW relationship is very similar to that of several forms.

Plots of HL versus HW (Fig. 13) and scape length (SL) versus HW (Fig. 21) are similar, but greater interspecific variability among the regressions for the SL–HW plot may offer more taxonomic potential. For example, despite greater intraspecific variability about the regression lines, more confident taxonomic judgements might be made using SL (Fig. 22) instead of HL (Fig. 14) with HW to separate N16 and N20 (but it makes little difference in differentiating N16

Figs. 19–24. Two-variable scattergrams with fitted regression lines.



Figs. 25–30. Two-variable scattergrams with fitted regression lines.



or N20 from N21b). The SL–HW relationship provides other useful and (or) interesting comparisons. Good separation of *fulva*, N21a (at least the northern dark ones), and N21b is obtained (Fig. 23). Separation of *fulva* from N20 and N22a (Fig. 24) is weaker, but, on average, *fulva* has comparatively shorter scapes for a fixed head width. Considerable overlap between the scatterplots of *fulva* and N16 (Fig. 25) and N22a and N22b (Fig. 26) means that the SL–HW relationship would not be very useful for separating each pair of forms when used alone, but may contribute to multivariate discrimination; this may also apply to the separation of N19 and N20 (Fig. 27). Highly coincident scatterplots of N17 and N22a (Fig. 28) indicate that the SL–HW relationship is of little value in separating these two forms. Note that N16 is fairly distinct from N22a and N22b (Fig. 26) and to a lesser degree from N17 and N18 (Fig. 29). Since *fulva* can be recognized by other means, the SL–HW relationship can separate N16 from the other eight forms.

Stepwise discriminant analysis selected interspinal length (ISPL) as the first variable to be entered into the model, while I have suggested that WLA is a good size variable. A plot of ISPL versus WLA (Fig. 31) indicates that interspecific and intraspecific variability in ISPL are both large for a fixed value of WLA (length of spines appears to have a positive allometric relationship with body size, particularly for *fulva*). On average, propodeal spines are longest in *fulva* and shortest in N21b for a fixed value of WLA, but the overlap of scatterplots between *fulva* and other forms (particularly N17, N18, and N21a) is considerable. Separation of N21b is reasonably good, but is improved by using tibial length (TL) instead of WLA on the horizontal axis (Fig. 32). This uses the fact that N21b has long legs, especially compared with other forms with relatively short propodeal spines.

The ISPL–WLA relationship gives excellent separation of N21b from N19, N20, and N21a, and appears to give fairly good separation of N21a from N20 and N19 (Fig. 33). However, at least part of the problem in classifying "N21a" test specimens using the discriminant functions seems to be attributable to relatively shorter propodeal spines in some N21a test colonies than in the regular data set. For example, the measured worker from one Big Pine Key colony had ISPL = 0.13, WLA = 1.64; using the estimated regression function of ISPL on WLA (Table 10) would give a predicted ISPL of 0.20. Of the four forms of the *rudis–texana* subcomplex that occur in Ontario, the darker N17 and N18 have longer spines, on average, than the lighter N16 and N22b (Fig. 34). For a fixed value of TL, N22a has shorter spines, on average, than N17 and N18, but there is still considerable overlap (Fig. 35).

The relationship between the width and length of the (hind) femur (Fig. 36) was surprisingly useful. The unusually thin femora of *fulva* are distinctive. Broader headed forms tend to have comparatively shorter and thicker femora, but *fulva* falls completely out of this pattern. Hence, *fulva* and N16, which share the same HL–HW relationship (Fig. 16) and overlap considerably in their SL–HW relationship (Fig. 25), can be separated easily by their relative femur dimensions (Fig. 37). The femora of N21b are relatively thin compared with those of several other forms, but their outstanding feature is their great length in larger individuals (Fig. 36).

Creighton (1950) used a comparatively larger eye size as one character that would separate *miamiana* (N21a) from *rudis* and *picea*. He measured eye size by the number of facets in the greatest diameter, whereas I have used eye length (EL), since the facets can be difficult to count. The large eye size of N21a is fairly noticeable, but as can be seen on a plot of EL versus WLA (Fig. 38), there is some overlap with other forms, particularly N20 (also see Fig. 41; two points looking like solid circles are overlapped symbols for N20 and N21a). Creighton's key had separated *carolinensis* (N20) from *miamiana* on the basis of a comparatively narrower head in the former, but I have shown here that there is little if any difference in the HL–HW relationship between these two forms. Undoubtedly EL (or some other measure of eye size) is a useful character for separating N21a from several other forms. For example, the EL–WLA relationship provides fairly good (although not completely unambiguous) separation of N21a from the other very dark forms, N17 and N18 (Fig. 39), in the rather unlikely event it was ever needed. Using HW rather than WLA on the x axis improves separation (Fig. 40). In a plot of EL on WLA, separation of N21a and N21b is quite distinct (Fig. 41).

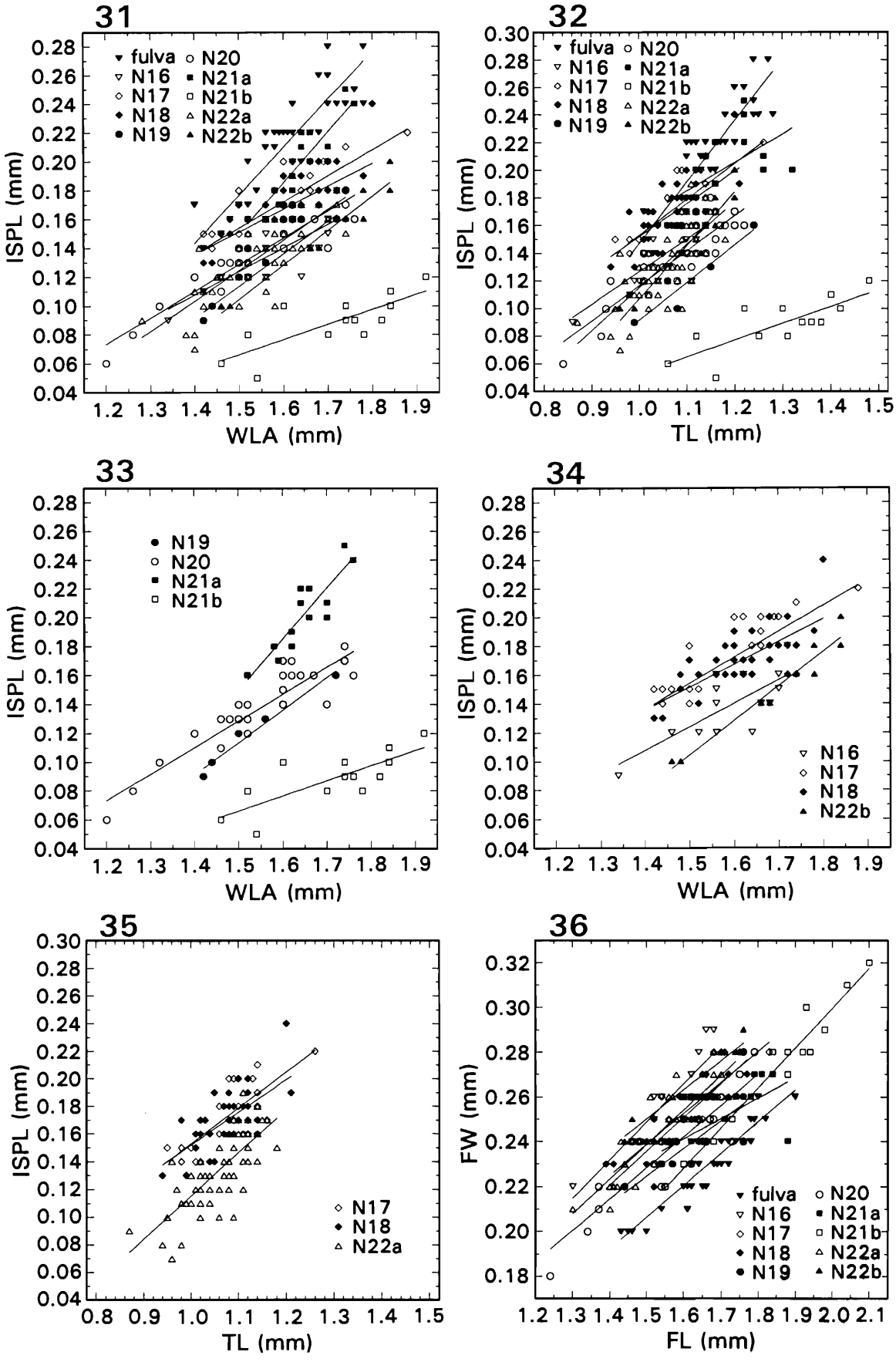
Figures 42–48 illustrate selected plots involving indices. If an index is independent of size, then a regression of the index on a size-dependent variable (WLA was used here) should result in regression lines that are parallel (within the bounds of sampling error) to the x axis. The cephalic index (CI) (Fig. 42) and scape index 1 (SI1) (Fig. 43) are highly correlated with size, as expected. The spinal index (SPI) also appears to be correlated with size (Fig. 44). The eye (or ocular) index (EI) appears to be somewhat negatively correlated with size (Fig. 45), but femur index 1 (FI1) appears to be relatively uncorrelated with size (Fig. 46). A plot of EI versus FI1 (Fig. 47) shows that some forms could be separated on the basis of the indices used, such as *fulva*, N16, and N22b from N21a using EI, and *fulva* from N16 and N22b using FI1. A plot of SPI versus FI1 (Fig. 48) also shows some separation of forms. However, the failure of these plots to clearly distinguish *fulva* from all other forms is a clear example of the inferiority of this method to the multivariate approach. It is also difficult to see what any of the indices offer in taxonomically useful information that is not more simply available from the two-variable scatterplots, especially given the close morphological similarity among the forms. The empirical evidence suggests that indices are of little value in the taxonomy of this complex, despite their obvious utility in other groups.

Discussion

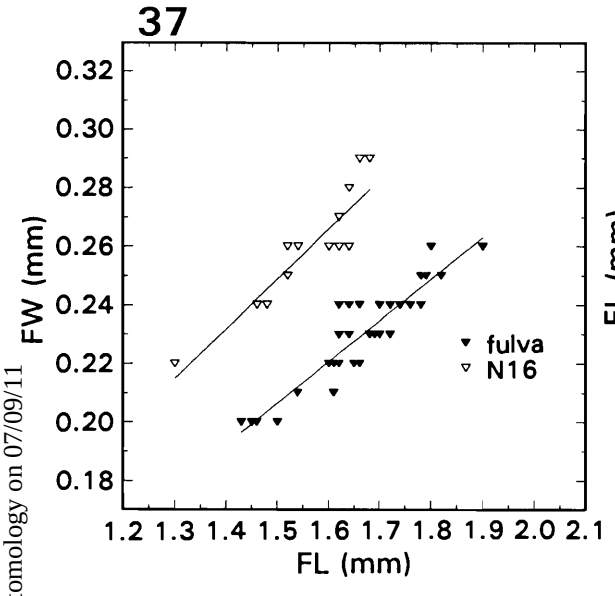
The species (or forms) recognized in this paper were first delineated by genetic markers that exhibit very little variability. Morphological differentiation among forms was then examined, primarily using morphometric methods; morphology (including morphometrics) was *not* used to initially define the forms. The multivariate analysis and scatterplots provide a morphometric description of "habitus" differences that accounts for size; even most minims can be identified with reasonable confidence. These morphometric analyses on the genetically identified colonies have provided a basis for judging the taxonomic status of the names *punctaticeps* and *crinimera*, although no live colonies attributable to these names were seen in this study.

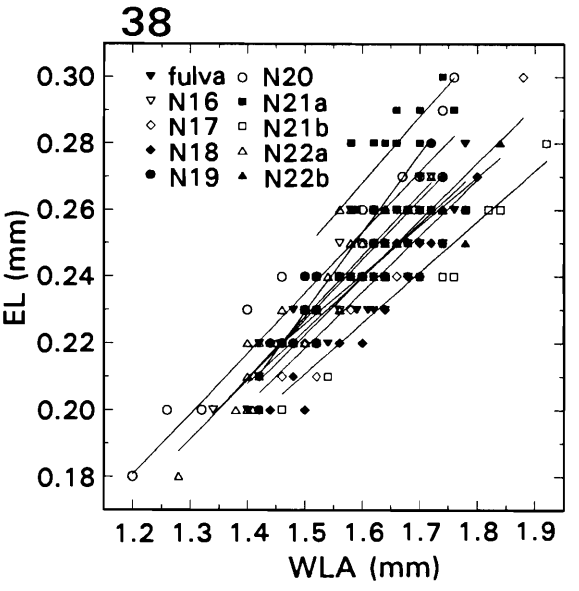
Errors in classification are inevitable with the 12 variables

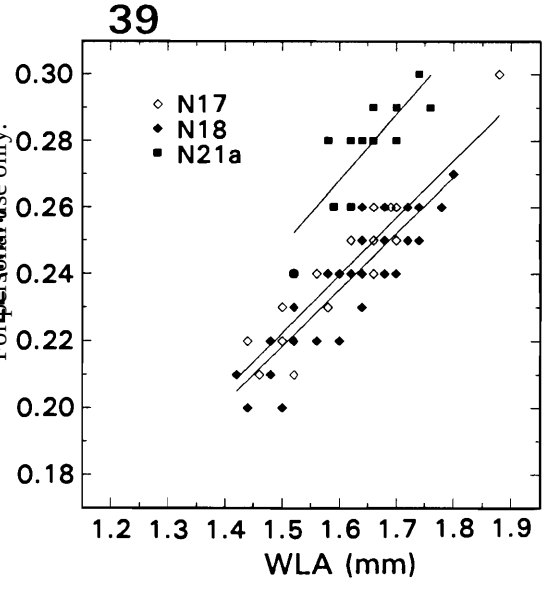
Figs. 31–36. Two-variable scattergrams with fitted regression lines.

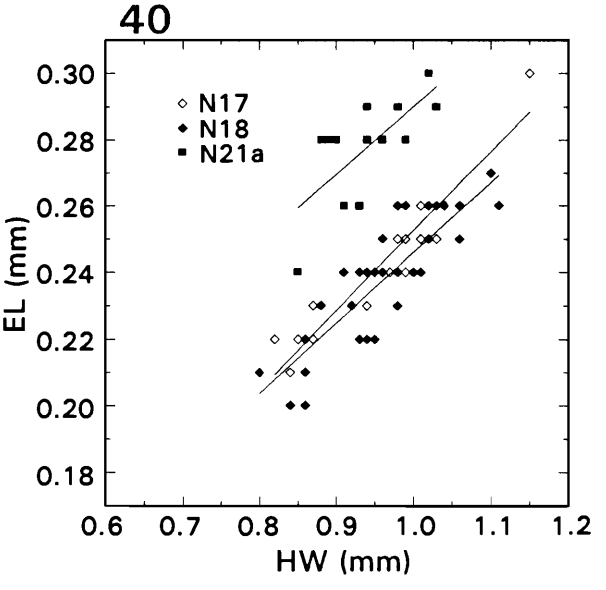


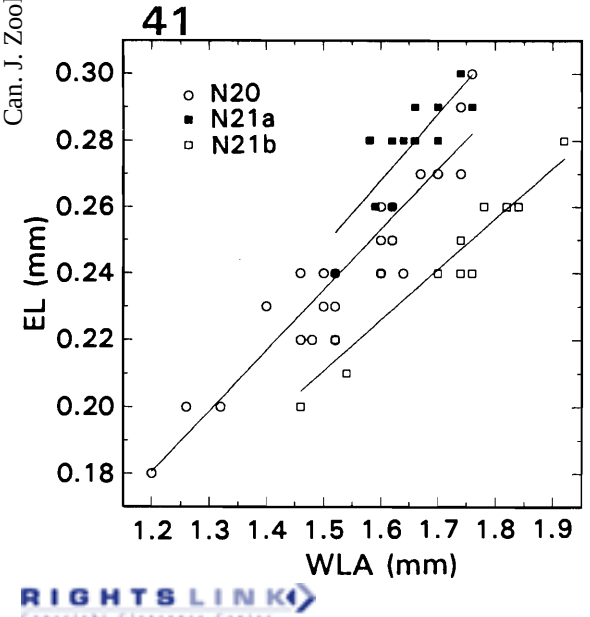
Figs. 37–42. Two-variable scattergrams with fitted regression lines.

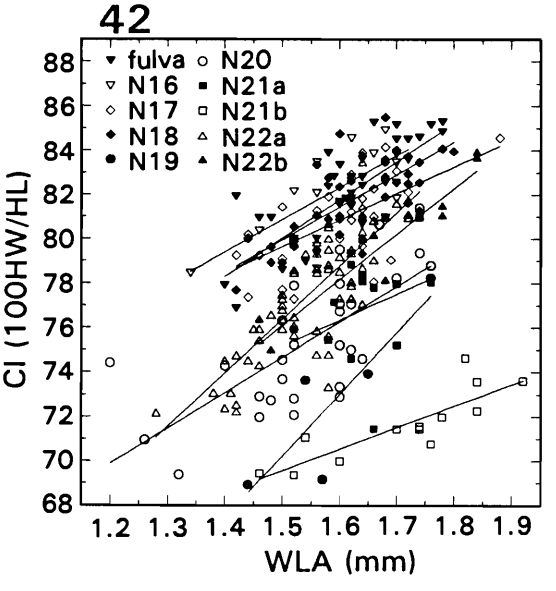




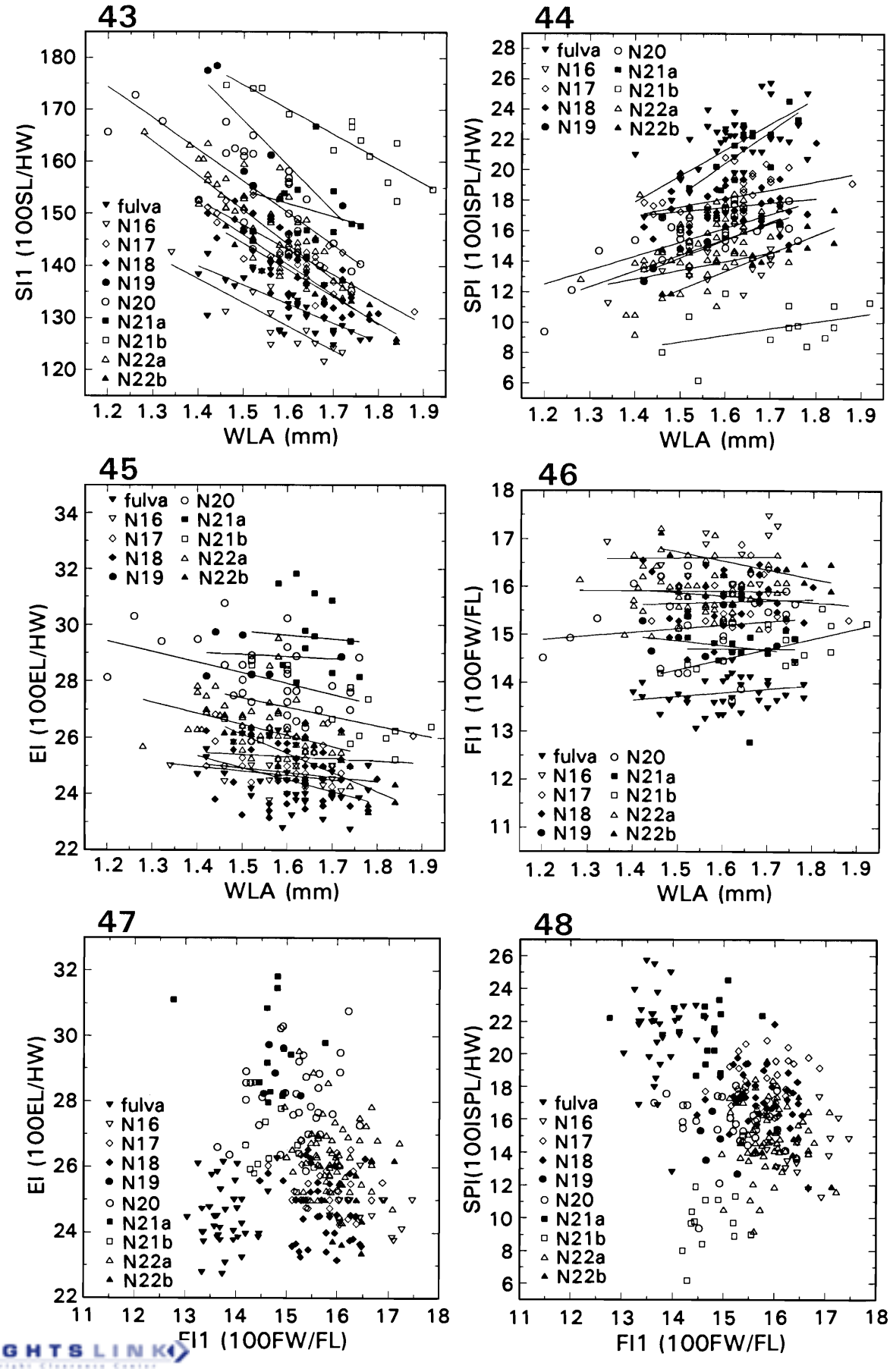








Figs. 43–48. Two-variable scattergrams, some with fitted regression lines.



used because of insufficient phenetic distance between some forms. The most serious errors occur between N17 and N18, both dark ants found in northeastern North America that appear to differ in karyotype by a single centric fission or fusion (G.J. Umphrey, unpublished data); these appear to be parapatric forms that require further testing for evidence of hybridization and introgression. However, on a local level sympatric forms are often differentiable, and some forms can be recognized confidently on a broad geographical basis. Used judiciously, colour can effectively discriminate between many pairs of forms, especially if a good reference collection is available. The key following summarizes the most important taxonomic characters recognized.

The multivariate classification might be improved by (*i*) using additional variables, (*ii*) using better measuring equipment, or (*iii*) measuring more ants per colony. The last is most likely to lead to improved results. Daly and Balling (1978) obtained better results for identifying Africanized honey bees by using the mean measurements of 10 bees per colony rather than individual bees. Better measuring equipment, particularly for positioning the specimen, would likely improve the results for individual specimens. Both EL and FW are very coarsely measured at a resolution of 0.01 mm. I see few possibilities for the addition of other useful variables, but they may exist.

Further sampling is not likely to affect overall results, but it would increase the reliability of the functions for taxonomic discrimination. While some forms have been diversely sampled, others have a small sample size or restricted geographical coverage. For example, the 38 colonies of *fulva* came from 22 localities broadly spread over 11 states, and the 67 colonies of N22a came from 23 localities in 11 states. In contrast, all but one of the N21b colonies used in the regular data set came from two localities in Arizona, and all five N19 colonies came from two localities in Missouri. The sampling for N21a also appears deficient for accounting for some heterogeneity outside of the localities sampled. Further sampling is required, especially through the midwest, and more intensive sampling is desirable within the range covered, particularly within Florida, with the specific objective of expanding the data sets of N19, N21a, and N21b. This is a very direct extension of the work to date that simply requires some time and resources.

Clusters of sibling species appear to be quite common in ants (Crozier 1981), and this complex appears to contain one of the largest such clusters. Since sibling species present special problems in taxonomic delineation and recognition, it is hoped that the methodology employed here can usefully be applied to other taxa.

A preliminary key to the workers of the *fulva–rudis–texana* complex

This key should be used when couplet 18 is reached in Creighton's (1950, p. 140) key to the species of *Aphaenogaster*. This includes *A. punctaticeps*, which seems to be restricted (north of Mexico) to the desert regions of southern New Mexico, Arizona, and Texas. As its karyotype is unknown, *punctaticeps* is brought out in the first couplet, but it is most similar in morphology to N21b (*texana*). The key is intended for North America north of Mexico, but especially eastern North America. Distributional data should be interpreted with care pending additional sampling.

As much as possible, simple morphological and distributional characters are used, but more confident diagnosis often requires using canonical variates, karyotypes, and (or) electrophoretic information. For several forms only the karyotype is completely diagnostic; electrophoresis is required only to separate N22a and N22b. Karyotype information includes the basic haploid chromosome number, but supernumeraries can exist; the additional chromosome information should provide a karyotypic diagnosis anyhow. The second couplet uses, in part, the sculpture of the queen's mesopleurae (anepisternum and katepisternum), as this is the best morphological character found in this complex; all other morphological comparisons are for workers.

Comparisons based on canonical variates require all 12 variables be measured as defined in the Materials and methods. Natural logarithms of these values are multiplied by the appropriate set of corresponding *raw* canonical coefficients in Tables 6 and 7; these products and the constant are added to give the canonical variate value. All measurements are given in millimetres.

Body colour descriptions refer especially to the head and alitrunk, and usually also to the petiole plus postpetiole. All except dark brown ants usually have somewhat darker gasters and often some infuscation on the head. Legs are usually lighter.

1. Legs exceptionally long for body size, usually with TL > 1.50 (likely not true for minims, but adjusting for head size, usually with TL > 0.28 + 1.2HW) (rare; restricted to desert regions of southern New Mexico, Arizona, and Texas; reddish brown; cephalic sculpture delicate for size) ..... *punctaticeps*
— Legs not exceptionally long for body size, almost always with TL < 1.50 (adjusting for head size, usually with TL < 0.28 + 1.2HW) 2

2. Queen with mesopleurae entirely rugose (Fig. 5*a*). Worker with scapes relatively short for HW (Figs. 21, 23, 24) and hind femora relatively long and thin (Figs. 36, 37), so that SL < 0.55 + 0.84HW and FW < 0.014 + 0.141FL; sculpture generally very heavy on medium-sized and larger workers, the forecoxa often rugose (Figs. 3*a*, 3*b*); propodeal spines relatively long (Figs. 31, 32) and often directed upwards; mesonotum often with anterior edge rising abruptly above the adjacent portion of the pronotum (Fig. 3*b*; mainly in medium-sized and larger workers); CAN1 < −3 and CAN2 > 0, using the coefficients of Table 6*a* (Fig. 6*a*); karyotype of $n = 18$, of which none of the six largest chromosomes are acrocentrics with very small short arms (widespread and abundant in eastern North America; fairly uniform light to dark reddish brown) ..... *fulva*
— Queen with mesopleurae largely smooth and shining (Fig. 5*b*). Worker with scapes relatively long for HW or hind femora relatively short and thick, so that SL > 0.55 + 0.84HW and (or) FW > 0.14 + 0.141FL (with rare exceptions); sculpture variable, but not as heavy as found on many *fulva* workers, the forecoxa with little sculpture (Figs. 3*d*–3*h*, 4*a*–4*h*); propodeal spines often relatively shorter and often directed backwards; mesonotum only occasionally (primarily in N17 and *picea*) with anterior edge rising abruptly above the adjacent portion of the pronotum, and never as strongly as found in many *fulva*; CAN1 > −3 or CAN2 < 0, using the coefficients of Table 6*a*; karyotype variable, but if $n = 18$, at least one of the six largest chromosomes (usually two of the largest three) are acrocentrics with very small short arms 3

3. Propodeal spines relatively short and legs relatively long, so that ISPL < −0.17 + 0.233TL (Fig. 32); CAN1 > 5 (and CAN2 > 2) using the coefficients of Table 6*a* (Fig. 6*a*); karyotype of $n = 21$, with no large metacentrics and no very small acrocentric present; range includes Texas to the mountains of southern Arizona but excludes eastern North America (colour widely variable, from yellow to dark brown) ..... N21b (*texana*)
— Propodeal spines relatively longer or legs relatively shorter, so that ISPL > −0.17 + 0.233TL (Fig. 32); CAN1 < 5 using the coefficients of Table 6*a* (Fig. 6*a*); karyotype of $n = 16-22$, with at least one large metacentric or (for $n = 22$ forms) a very small acrocentric present; range includes eastern North America but likely excludes the southwestern United States from western Texas to the mountains of southern Arizona 4

4. Scapes short relative to HW, so that SL < 0.632 + 0.687HW (Fig. 30, also see Figs. 26, 29); karyotype of $n = 16$, of which five of the seven largest chromosomes are metacentrics subequal in size (medium brown, sometimes a bit darker, head and especially gaster often darker than alitrunk; known from Missouri to New Jersey and north to southeastern Canada) N16
— Scapes longer relative to HW, such that SL > 0.632 + 0.687HW (rare exceptions); karyotype of $n \geq 17$, of which no more than four of the seven largest chromosomes are metacentrics 5

5. Dark brown (piceous), with eyes relatively small for HW (EL < −0.017 + 0.255HW), and occurring in northeastern North America and throughout higher elevations of the Appalachians; CAN1, CAN2 falling within zone for N17 and N18 on Fig. 7, using coefficients of Table 6*b*; karyotype of $n = 17$ or 18. (These forms are most likely to be confused within their range with the medium-brown N22a, from which they can be separated in part by using the plot of ISPL on TL in Fig. 35. Better separation is obtained by calculating CAN1 from the coefficients in Table 6*b*, and comparing this value with the plot in Fig. 10) 6
— Light reddish brown to medium brown, except for N21a, which is dark brown outside of southern Florida but has relatively large eyes for HW (EL > −0.017 + 0.255HW for N21a, variable otherwise; Fig. 40) and occurs in southeastern United States at lower elevations (restricted to coastal plain?) outside of range of N17 and N18; CAN1, CAN2 falling outside zone for N17 and N18 in Fig. 7, using coefficients in Table 6*b*; karyotype of $n \geq 19$ 7

6. Karyotype of $n = 17$, with one metacentric substantially larger than the other chromosomes; range includes Ontario, Ohio, and Kentucky, but is probably mainly to the west of range of N18, and likely excludes the higher elevations of the Appalachians. (The forms in this couplet are essentially indistinguishable morphologically.) N17
— Karyotype of $n = 18$, with the largest metacentric subequal in size to two acrocentrics; range includes the higher elevations of the Appalachians, Ontario, Connecticut, New York, and Pennsylvania, and likely several other eastern states and provinces, mainly east of the range of N17 ..... N18 (*picea*)

7. Karyotype of $n = 19-21$, with one or two very large metacentrics; CAN1, CAN2 falling inside zone (defined approximately as CAN1 > 1) for N19, N20, and N21a in Fig. 7, using coefficients in Table 6*b*; at least one form present throughout Florida and coastal plain of the southeastern United States, but not confined there 8
— Karyotype of $n = 22$, with no large metacentric; CAN1, CAN2 falling inside zone (defined approximately as CAN1 < 1) for N22a and N22b in Fig. 7, using coefficients in Table 6*b*; both forms apparently absent from all of Florida and the coastal plain of the southeastern United States. (The most difficult forms to separate at this couplet are N20 and N22a, which are found sympatrically in the piedmont of the southeastern United States. Using the coefficients in Table 7*c* to obtain CAN1 and CAN2 values, Fig. 11 will aid in separating these two forms. As well, N22a tends to be medium brown, while N20 tends to be light reddish brown, but only the karyotypes are diagnostic.) 10

8. Karyotype of $n = 21$, with only one very large metacentric; propodeal spines often relatively long (Fig. 33); CAN1, CAN2 falling inside zone for N21a in Fig. 7, using coefficients in Table 6*b*; dark brown (piceous) outside of southern Florida, where they are a lighter reddish brown (range probably extends throughout the coastal plain from North Carolina to eastern Texas and includes all of Florida) ..... N21a (*miamiana*)
— Karyotype of $n = 19$ or 20, with two very large metacentrics; propodeal spines usually relatively short (Fig. 33); CAN1, CAN2 falling inside zone for N19 and N20 in Fig. 7, using coefficients in Table 6*b*; yellowish brown to reddish brown but never dark brown, and likely absent from southern Florida 9

9. Karyotype of $n = 20$, very small acrocentric present; known range includes the piedmont of the southeastern United States and the whole of the coastal plain from southern New Jersey to the Florida panhandle ..... N20 (*carolinensis*)
— Karyotype of $n = 19$, very small acrocentric absent; known only from Missouri (probably found throughout much of the Mississippi basin) N19

10. Larger workers frequently with HW > 1.05 mm (Fig. 26); CAN1, CAN2 falling within N22b zone in Fig. 9, using coefficients in Table 7*a*; unique fast allele for isocitrate dehydrogenase present; scarce (light to medium reddish brown; known from Missouri to Maryland and north to southern Ontario) N22b
— Larger workers rarely with HW > 1.05 mm; CAN1, CAN2 falling within N22a zone in Fig. 9, using coefficients in Table 7*a*; common slow allele for isocitrate dehydrogenase present; very common and widespread (generally medium brown, only occasionally approaching lighter colour of N22b or darker colour of N17 and N18; known range includes Missouri, Alabama, Georgia, Indiana, and New Jersey, but it has not been found in Canada or Florida and may be absent from most of the coastal plain of the southeastern United States) ..... N22a (*rudis*)

Acknowledgments

I thank L. Ling for the preparation of the scanning electron
RIGHTSLINK
graphs, J. Swann for preparing critical-point-dried males, and S. Marshall for the use of his laboratory for making the measurements. I thank the following for field or laboratory assistance or the gift or loan of live colonies or voucher specimens: J. Carroll, J. Corrigan, S. Cover,

R. Crozier, J. Cumming, L. Ling, K. McLachlin-Hamilton, S. Porter, R. Snelling, J. Trager, J. Umphrey, and D. Wheeler. D. Ankney, J. Helava, P. Jaspers-Fayer, and I. MacNeill provided substantial computing assistance. Financial assistance for this work was provided by the Natural Sciences and Engineering Research Council of Canada through a postgraduate scholarship to the author and a research grant to H.F. Howden, who also provided support in many other forms. The suggestions of the reviewers are most appreciated.

References

Atchley, W.R., Gaskins, C.T., and Anderson, D. 1976. Statistical properties of ratios. I. Empirical results. Syst. Zool. **25**: 137–148.

Baker, R.J., Atchley, W.R., and McDaniel, V.R. 1972. Karyology and morphometrics of Peters' tent-making bat, *Uroderma bilobatum* Peters (Chiroptera, Phyllostomatidae). Syst. Zool. **21**: 414–429.

Blackith, R.E., and Reyment, R.A. 1971. Multivariate morphometrics. Academic Press, London.

Cagniant, H. 1989. Contribution à la connaissance des fourmis marocaines. Description des trois castes d'*Aphaenogaster weulersseae* n.sp.; notes biologiques et écologiques; étude comparée de trois populations [Hym. Formicoidea Myrmicidae]. Bull. Soc. Entomol. Fr. **94**: 113–125.

Carroll, J.F. 1975. Biology and ecology of ants of the genus *Aphaenogaster* in Florida. Ph.D. thesis, University of Florida, Gainesville.

Cole, A.C., Jr. 1953. Studies of New Mexico ants. II. A description of a new subspecies of *Aphaenogaster huachucana* (Hymenoptera: Formicidae). J. Tenn. Acad. Sci. **28**: 82–84.

Creighton, W.S. 1950. The ants of North America. Bull. Mus. Comp. Zool. **104**.

Crozier, R.H. 1975. Hymenoptera: animal cytogenetics. Vol. 3. Insecta 7. Gebrüder Bornträger, Berlin.

Crozier, R.H. 1977. Genetic differentiation between populations of the ant *Aphaenogaster* '*rudis*' in the southeastern United States. Genetica (The Hague), **47**: 17–36.

Crozier, R.H. 1981. Genetic aspects of ant evolution. *In* Evolution and speciation: essays in honour of M.J.D. White. *Edited by* W.R. Atchley and D. Woodruff. Cambridge University Press, Cambridge. pp. 356–370.

Crozier, R.H., Pamilo, P., Taylor, R.W., and Crozier, Y.C. 1986. Evolutionary patterns in some putative Australian species in the ant genus *Rhytidoponera*. Aust. J. Zool. **34**: 535–560.

Daly, H.V., and Balling, S.S. 1978. Identification of Africanized honeybees in the western hemisphere by discriminant analysis. J. Kans. Entomol. Soc. **51**: 857–869.

Elmes, G.W. 1978. A morphometric comparison of three closely related species of *Myrmica* (Formicidae), including a new species from England. Syst. Entomol. **3**: 131–145.

Emery, C. 1895. Beiträge zur Kenntniss der nordamerikanischen Ameisenfauna. Zool. Jahrb. Abt. Syst. Geogr. Biol. Tiere, **8**: 257–360.

Francoeur, A. 1973. Révision taxonomique des espèces néarctiques du groupe *fusca*, genre *Formica*. Mem. Soc. Entomol. Que. **3**: 1–316.

Lachenbruch, P.A. 1975. Discriminant analysis. Hafner Press, New York.

Lachenbruch, P.A. 1982. Discriminant analysis. *In* Encyclopedia of statistical sciences. Vol. 2. *Edited by* S. Kotz and N.L. Johnson. John Wiley & Sons, New York. pp. 389–397.

Lubischew, A.A. 1962. On the use of discriminant functions in taxonomy. Biometrics, **18**: 455–477.

MacKay, W.P. 1989. A new *Aphaenogaster* (Hymenoptera: Formicidae) from southern New Mexico. J. N.Y. Entomol. Soc. **97**: 47–49.

Menozzi, C. 1929. Una nuova specie di formica del genere *Aphaenogaster* Mayr del Nord America. Bull. Lab. Zool. Gen. Agr. Portici, **22**: 282–284.

Reyes Lopez, J.L., and Porras Castillo, A. 1984. Alar biometry in the taxonomy of the species *Goniomma hispanicum* and *G. baeticum*. Insectes Soc. **31**: 473–475.

SAS Institute Inc. 1988. SAS/STAT user's guide, release 6.03 ed. SAS Institute Inc., Cary, N.C.

Umphrey, G.J. 1992. Differentiation of sibling species in the ant genus *Aphaenogaster*: karyotypic, electrophoretic, and morphometric investigations of the *fulva–rudis–texana* complex. Ph.D. thesis, Carleton University, Ottawa, Ont.

Ward, P.S. 1980. A systematic revision of the *Rhytidoponera impressa* group (Hymenoptera: Formicidae) in Australia and New Guinea. Aust. J. Zool. **28**: 475–498.

Ward, P.S. 1985. The Nearctic species of the genus *Pseudomyrmex* (Hymenoptera: Formicidae). Quaest. Entomol. **21**: 209–246.

Ward, P.S. 1989. Systematic studies on pseudomyrmecine ants: revision of the *Pseudomyrmex oculatus* and *P. subtilissimus* species groups, with taxonomic comments on other species. Quaest. Entomol. **25**: 393–468.

Ward, P.S. 1993. Systematic studies on *Pseudomyrmex* acacia-ants (Hymenoptera: Formicidae: Pseudomyrmecinae). J. Hymenopt. Res. **2**: 117–168.

Wing, M.W. 1968. Taxonomic revision of the Nearctic genus *Acanthomyops* (Hymenoptera: Formicidae). Mem. N.Y. Agric. Exp. Stn. (Ithaca), No. 405.