Tanzania Journal of Science 44(1): 37-51, 2018 ISSN 0856-1761, e-ISSN 2507-7961

THE POWER OF COEFFICIENTS AND METHODS OF CODING IN DELIMITING SPECIES USING PHENETIC APPROACH: THE CASE OF AFRICAN *SOLANUM* SECTION *SOLANUM SENSU* EDMONDS

Mkabwa LK Manoko
University of Dar es Salaam, College of Agricultural Sciences and Fisheries Technology
Department of Crop Sciences and Beekeeping Technology, P.O. Box 60091, Dar es Salaam. Tanzania.
manoko@udsm.ac.tz

ABSTRACT

Phenetics is one of approaches used to delimit species in plant classification. Conclusion in phenetics is based on overall similarity often of morphological data. The approach uses coded data that are analyzed using coefficients to create similarity matrices that are analyzed using clustering analysis to create the classification. Exists different similarity coefficients and coding methods though in practice are used intuitively sometimes giving results that have been challenged. Though similarity coefficients and coding methods have some times been blamed, studies to analyze their influences are limited. The trend however is to avoid morphological data in favor of DNA markers. The current study assessed the power of eight similarity coefficients to recover ten known section Solanum species that have also been delimited using AFLPs. Each similarity coefficient was used to analyze two similarity matrices created using two Pledji's binary or conventional methods of coding multistate characteristics. Analysis used clustering option of PAST's software. The ten species were recovered from each matrix only when Gower's or Hamming's coefficients were used. Jaccard's and Dice coefficients recovered ten species only with binary coding. Other coefficients recovered zero to five species. Coefficients of similarity and coding method thus influences species level classification in phenetic approach.

Key words: Phenetic approach, similarity coefficients, UPGMA, Section *Solanum*, cophenetic coefficient, morphology.

INTRODUCTION

Botanical classification remains to be one of the most important undertaking that serve mankind. It finds its use in all sciences and traditions that use plants though often underestimated. The basic unit of classification is the species based on which all other more inclusive ranks are formed. Proper delimitation of species is key to understanding of biodiversity and develop strategy for the conservation. Proper identification which can only be done based on descriptions of species have saved humanity. BioNET lists 48 case studies where correct identification of species was the key to success. The notable areas were epidemiology, pest management and quarantine, human health, pathogens and their biological control agents, protection of useful plants and insects (http://www.bionet-intl.org/opencms/opencms/caseStudies/default.jsp). Species identification is only possible if the boundaries between and among species of the same genus are clear.

Traditionally morphological characteristics have been used to delimit species though recently use of DNA markers is at increase (Hebert et al. 2003, Hebert et al. 2004, Pons et al. 2006, Clement and Donoghue 2012). This trend however has led to a debate on the suitability of DNA markers over

morphology and vice versa (Hebert et al. 2004, Brower 2006, Sass et al. 2007, Taberlet et al. 2007, Spooner, 2009, Liu et al. 2010). Although this debate is beyond the scope of this study, it is necessary to mention here that in plant classification morphological characteristics are still extremely important. All taxa are practically identified, descriptions are written and keys are constructed using morphology. According to the International Code of Botanical Nomenclature a plant species has only one correct name. A botanical name is given to a group of individuals that can be described using morphology and is different from other related groups. Thus even when DNA markers have been used, the practice is to correlate the groups defined by molecular markers to morphology (Bohs and Olmstead 1997, Bohs 2005, Levin et al. 2005, Levin et al. 2006, Weese and Bohs 2007). Morphological data also allows the interpretation of the observed features of plants and the formation of hypotheses on adaptation and evolution (Knapp 2001).

Phenetic taxonomy is a system of classification based on the overall similarity of the organisms being classified in which the relationship is based on all available data characters without any weighting (Sokal 1986). At each level members of each taxon are on the average more similar to each other than they are to members of other taxon at corresponding levels. At species level classification therefore individuals that fall in the same cluster constitute the same species whereas those that exhibit morphological or genetic gaps they fall in different clusters and they are thus different species. Clustering pattern in phenetic approach and even in phylogenetic reconstruction, is influenced by among others method of coding of multistate characteristic (Sokal 1986, Jackson et al. 1989, Wiens 2000, Datwyler and Wolfe 2004, Simmons and Geisler 2002). Pledji's (in Forey and Kitching 2000) described four methods of coding. It has also been said that, phenetic results are influenced by the subjective choice of coefficient of similarity (Jackson et al. 1989, Sokal 1986, Finch 2005). Nevertheless, though there are many coefficients, only a few are commonly used intuitively. The frequently used coefficients are Simple Matching, Jaccard's and Euclidean (Sokal 1986, Finch 2005). Jackson et al. (1989) compared six similarity coefficients namely Jaccard's, Dice, Rusell and Rao, Simple Matching, Rogers-Tanimoto, Ochiai, Phi and Yule. These authors concluded that the dendrograms obtained provided little evidence of group structure and some coefficients provided more or less similar information. Edmonds (1978) studied member of the sect. *Solanum* using phenetic approach. The author found that nine different sets of morphological data sampled from the same individuals depicted different classifications. Similarly, Olet (2004) failed to separate eight known species of sect. *Solanum* from each other. Whether or not this confusion could be explained by the used similarity coefficients or method of coding has never been assessed. Sokal (1986) called for studies to compare usefulness of different coefficients of similarity but such studies are limited. What is however evident currently is avoidance to use morphological data in classification in favor of DNA markers.

Section *Solanum* is one of the most taxonomically complex groups in the genus *Solanum* when it comes to species delimitation. The complication is attributed to existence of genetically determined variations coupled with environmentally induced phenotypic plasticity, existence of different ploidy levels and polymorphism. Others are occurrence of natural hybridization between certain diploid taxa with various stages of pre- and post-fertilization isolating mechanisms (Edmonds and Chweya 1997). Current taxonomic treatment pulls together sections *Solanum*

Episarcophyllum, *Campanulisolanum*, and *Parasolanum* into one Morelloid clade (Bohs 2005, Bohs and Wiens 2007).

Section *Solanum* forms one of the most important groups of leafy vegetable in Africa. Members of the section are known for their medicinal, mollucidal or larvicidal properties and some carry resistance genes against *Phytophthora infestans* an important disease for cultivated Solanaceous crops such as Tomato and Irish potato (Roddick, 1991, Edmonds and Chweya1997, Kamoun et al. 1999, Singh et al. 2001, Zengfu, 2001, Heo et al. 2005). *Solanum nigrum* is poisonous and hyperaccumulates heavy metals (Perez et al. 1998, Xu et al. 2009). Such a taxonomically complicated but economically useful group makes a good candidate to assess of delimitation of species using morphology.

This study answers the following questions: (1) how does coefficient of similarity influences clustering pattern thus delimitation of species under phenetic approach? (2) Do method of coding multistate characteristics matter in phenetics analysis? (3) Whether or not phenetic classifications based on morphology lead to grouping unrelated forms into paraphyletic or even polyphyletic taxa making it possible to recognize multiple species? Usefulness of cophenetic coefficient measures was also assessed.

MATERIAL AND METHODS

Plant materials used in this study were grown at Radboud University Botanical garden from seeds obtained mainly from African countries. Table 1 summarizes seed accession numbers, the code used during the analysis and number of individuals per species. Morphological data were collected based on a descriptors list of 33 characteristics (both qualitative and quantitative) modified from Edmonds and Chweya (1997). Data were collected from plant of same age. Nomenclature was based on species recognized based on AFLP markers (Manoko 2007, Manoko et al. 2007, Olet et al. 2011, Manoko et al. 2012, Edmonds 2012).

Table 1: List of species

Botanical name	***Acronym***	***Individuals***
Solanum villosum Mill.	vill/VILL	24
Solanum nigrum L.	nigr/NIGR	12
Solanum nodiflorum Jacq.	nod/NOD	14
Solanum scabrum Mill.	scab/SCAB	31
Solanum chenopodioides Lam.	chen/CHEN	5
Solanum memphiticum Gmel. *sensu* Manoko 2007 non Edmonds 2007; 2012	mem/MEM	7
Solanum grossidentatum A. Rich. *Sensu* Manoko 2007 non Edmonds 2007; 2012	gross/GROSS	6
Solanum tarderemotum Bitt.	tar/TAR	8
Solanum umalilaense Manoko	uma/UMA	2
Solanum florulentum Bitt.	flor/FLOR	5

To assess the effect of coding of multistate qualitative characteristics on the resulting classification, two Pledji's methods of coding cited in Forey and Kitching (2000) were randomly selected. Each method was used to create a matrix. The selected methods were: (1) Method A in which multistate qualitative characteristics were coded as multistate (also called conventional coding). In this case, each characteristic

state was given a unique number i.e. 0, 1, 2 etc. Binary characters are coded as 0 or 1 and quantitative data are entered as continuous (2) Method D in which each character state of multistate character was considered as a variable and coded as 0 or 1 in an individual. Binary characteristics and quantitative characters were coded as in method A above. Latter in the text, these methods are referred to as conventional or binary, respectively.

All matrices were analyzed using PAST software Version 2.08 cluster program that performs UPGMA (Hammer et al. 2001). During the analysis eight coefficients of similarity were used. These are: Gower, Euclidean, Rho, Hamming, Manhattan, Jaccard and Dice. In addition, data in conventionally coded matrix were tagged according to their types i.e. binary, nominal or ordinal and analyzed using a mixture of coefficients option. This set includes coefficients that are used frequently such as Jaccard's and Euclidean and those designed for mixed data type sets such as Gower. Cluster analysis for each coefficient was analyzed using PAST default settings.

For each analysis cophenetic correlation coefficient scores were recorded and latter used as a measure of degree of fit of classification to the data set and also as a yard stick to choose best trees (Sokal et al. 1962, Farris 1969). The best dendrogram were those that reproduced the ten known good species that have been also recovered using AFLPs (Manoko 2007, Manoko et al. 2007, Olet et al. 2011, Manoko et al. 2012).

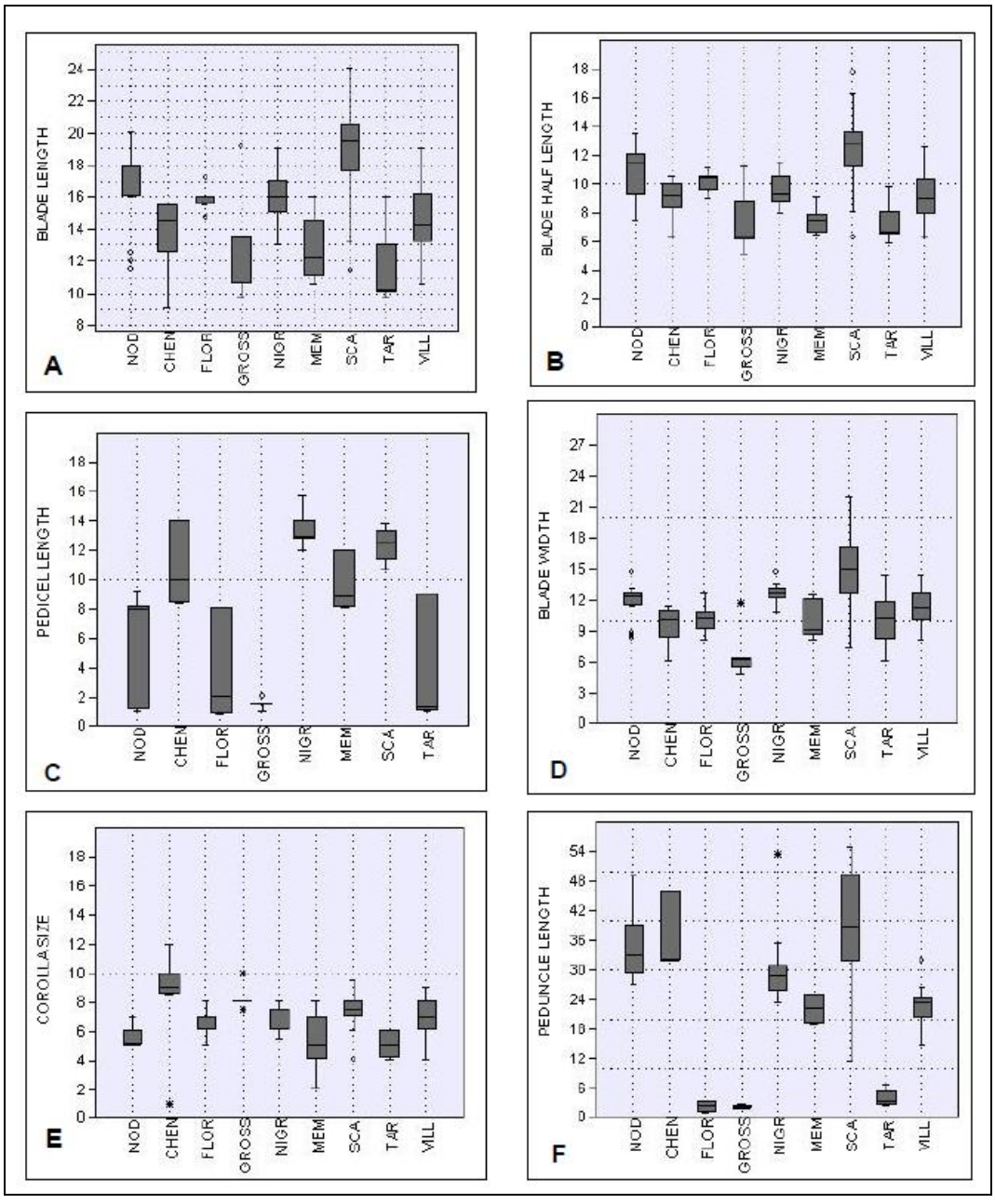
Box plots were used to identify outliers, which were *a posteriori* removed from some analyses. Box plot were also used to assess usefulness of each quantitative character based on rounding option as explained by Hammer et al. (2001). Resulting dendrograms are summarized in a table arranged based on their cophenetic correlation coefficient scores starting from the highest.

RESULTS

Assessment of usefulness of quantitative data using box plots

Figure 1 (A-I) present results on assessment of efficacy of the nine quantitative characteristics to differentiate the ten species studied. Based on these figures, quantitative characteristics studied can be grouped into 3 groups: Group 1 composed of characteristics that could split species at least in two groups regardless of inclusion of outliers. They included peduncle length, pedicel length, anther length, and style length. Group 2 composed of characteristics that could only differentiate species after outlier removal, which were blade length, blade width and corolla length. Group 3 was made up of characteristics that failed to differentiate species at all and this was made of only one character, the blade half-length.

Based on these results, eleven individuals from three species namely *S. scabrum, S. memphiticum* and *S. nodiflorum* were identified as outliers. These were 90023scab, 95115scab, 94125scab, 99023scab and 99012scab. Others were A1022mem, A1023mem, A1164bflorB, A3454nod, A3455nod and A3453nod.


BLADE LENGTH
A
NOD
CHEN
FLOR
GROSS
NIGR
MEM
SCA
TAR
VILL
BLADE HALF LENGTH
B
PEDICEL LENGTH
C
BLADE WIDTH
D
COROLLA SIZE
E
PEDUNCLE LENGTH
F

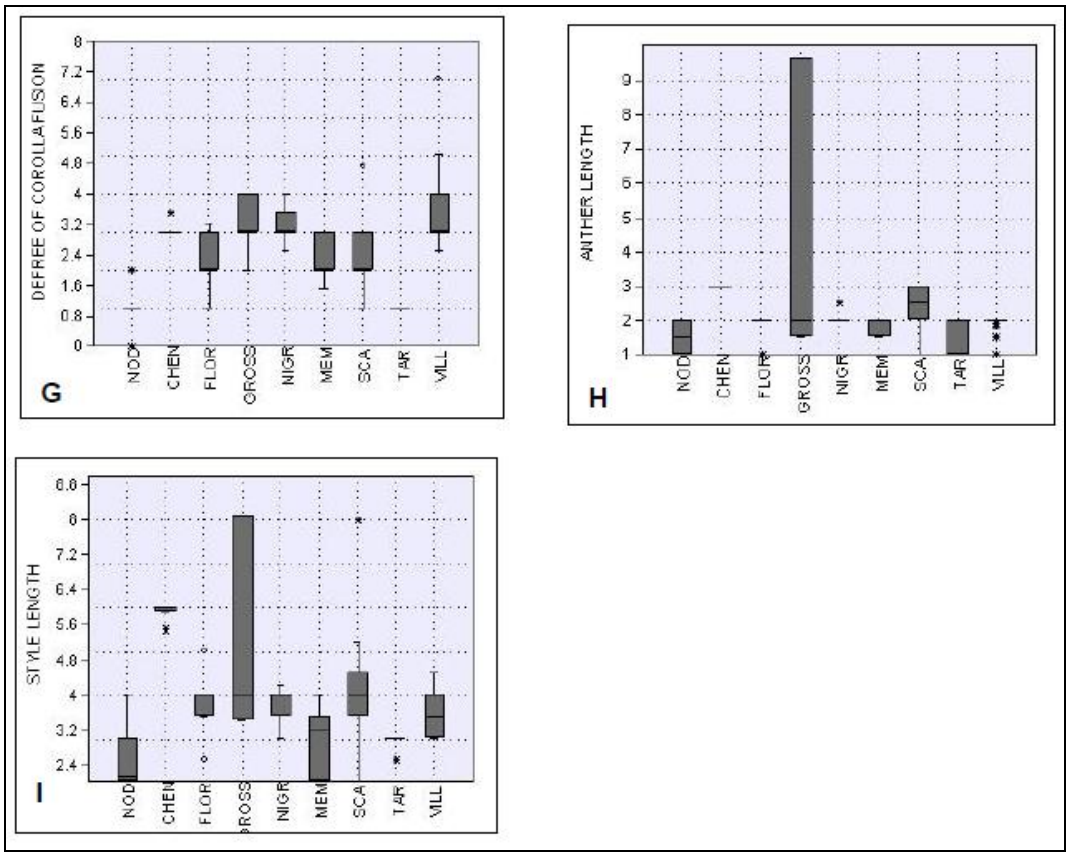


Figure 1: Box plots depicting variation of quantitative characteristics for each species. The bars are standard error of the mean. Stars represent outlier individuals.

Pattern of clustering

Table 2 presents the resulting dendrograms following the eight analyses performed.

Based on Table 2, only 6 of the fifteen dendrograms recovered all the ten species. Cophenetic correlation value for these dendrograms ranged from 0.7669 to 0.8013. Four of these dendrograms were those produced based on the conventionally or binary coded matrices analyzed using Hamming's or Gower's coefficients. Two were dendrograms produced based on binary coded matrix analyzed using Jaccard's or Dice coefficients of similarity. Of the remaining, though produced five to nine clusters, most of the clusters were not species specific. Eight dendrograms formed zero to three species specific clusters. These were all dendrograms analyzed with Manhattan and Euclidean, the one coded conventionally but analyzed using Rho coefficient and the one produced using a matrix created with mixed coefficients. Others were the two matrices that were conventionally coded but analyzed using Jaccard's or Dice coefficients. Though most of these had their cophenetic correlation value below 0.7669, cophenetic values of clusters formed by mixed coefficients and binary coded analyzed with Rho coefficient were compared to the former. Actually, a dendrogram obtained from conventionally coded matrix analyzed with Rho coefficient recorded the highest cophenetic correlation

value. A dendrogram resulting from the analysis of binary coded data analyzed using Rho coefficient produced five species specific clusters. Figure 2 present one of the six dendrograms that recovered all the ten species. On the other hand, figure 3 presents one of the dendrograms that produced between none to three species specific clusters. This is based on a conventionally coded matrix analyzed using Rho coefficient. It is a dendrogram that exhibited the highest cophenetic correlation value but recovered only two species. Figure 4 presents a pattern of clustering with inclusion of outliers. In this dendrogram the known species were recovered and outliers spread on different parts of the dendrogram.

Table 2: A summary of the clustering patterns of the all fifteen dendrograms obtained (Dendrograms that recovered ten species are in bold)

Coding	***Coefficient***	***Cophenetic value***	***Dendrograms' description***
Conventional	Rho	0.8031	Seven clusters only two species specific.
Conventional	**Hamming**	0.8013	Ten clusters formed each species specific.
Binary	**Jaccard**	0.7998	Ten clusters formed each species specific.
Binary	**Gower**	0.7941	Ten clusters formed each species specific.
Conventional	**Gower**	0.7953	Ten clusters formed each species specific.
Binary	Rho	0.7836	Eight clusters formed five species specific.
Binary	**Hamming**	0.7825	Ten clusters formed each species specific.
Conventional	Mixed	0.7713	Six clusters formed only two species specific.
Binary	**Dice**	0.7669	Ten clusters formed all species specific.
Binary	Manhattan	0.7063	Six clusters formed only one species specific.
Conventional	Jaccard	0.6954	Nine clusters formed three species specific.
Conventional	Manhattan	0.6926	Ten clusters formed none species specific.
Conventional	Euclidean	0.6923	Seven clusters formed, one species specific.
Binary	Euclidean	0.6844	Five clusters formed none species specific.
Conventional	Dice	0.6806	Eight clusters formed, one species specific.

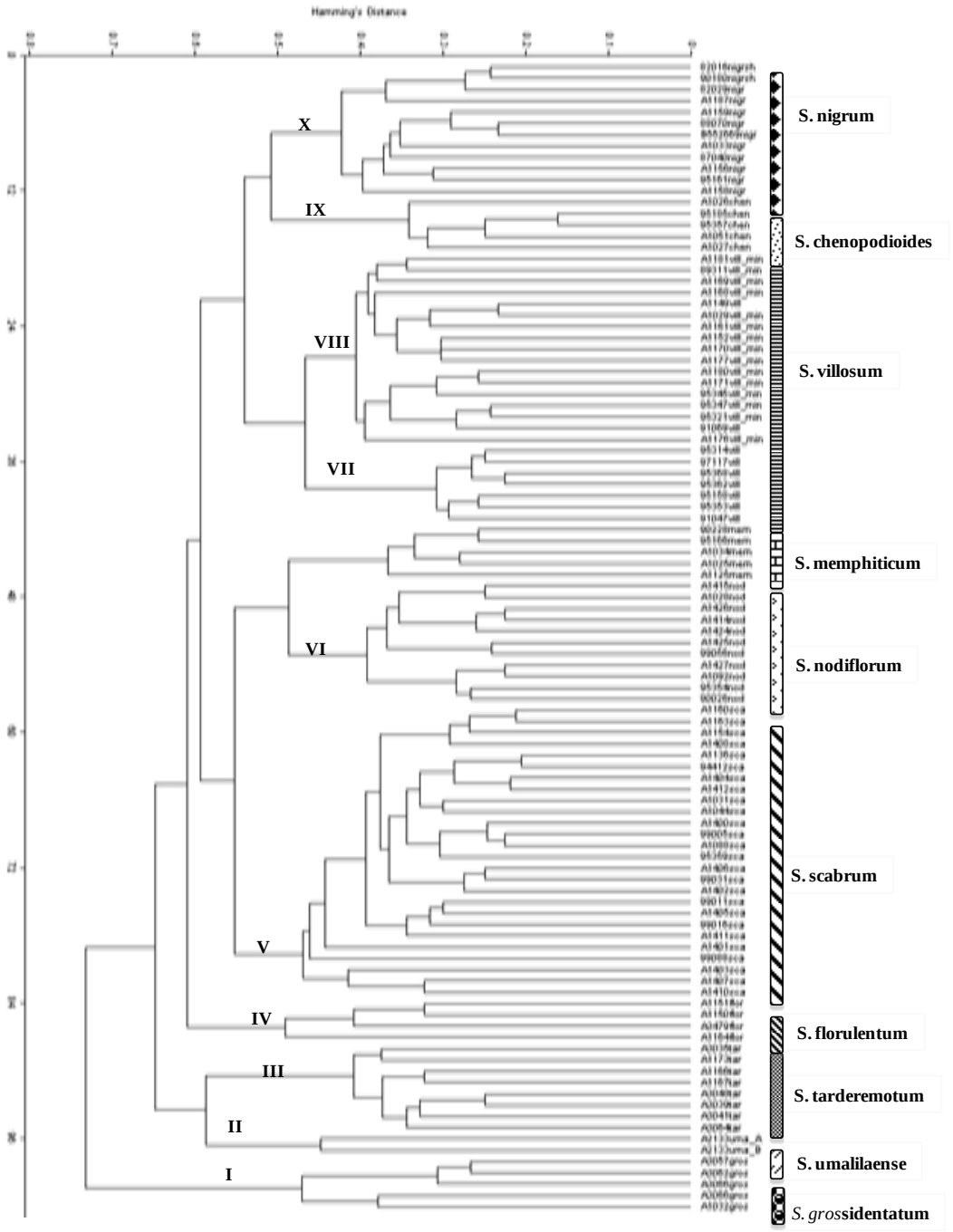


Figure 2: Dendrograms obtained using Gower's coefficient presenting the ten clusters labelled I - X each representing one known species

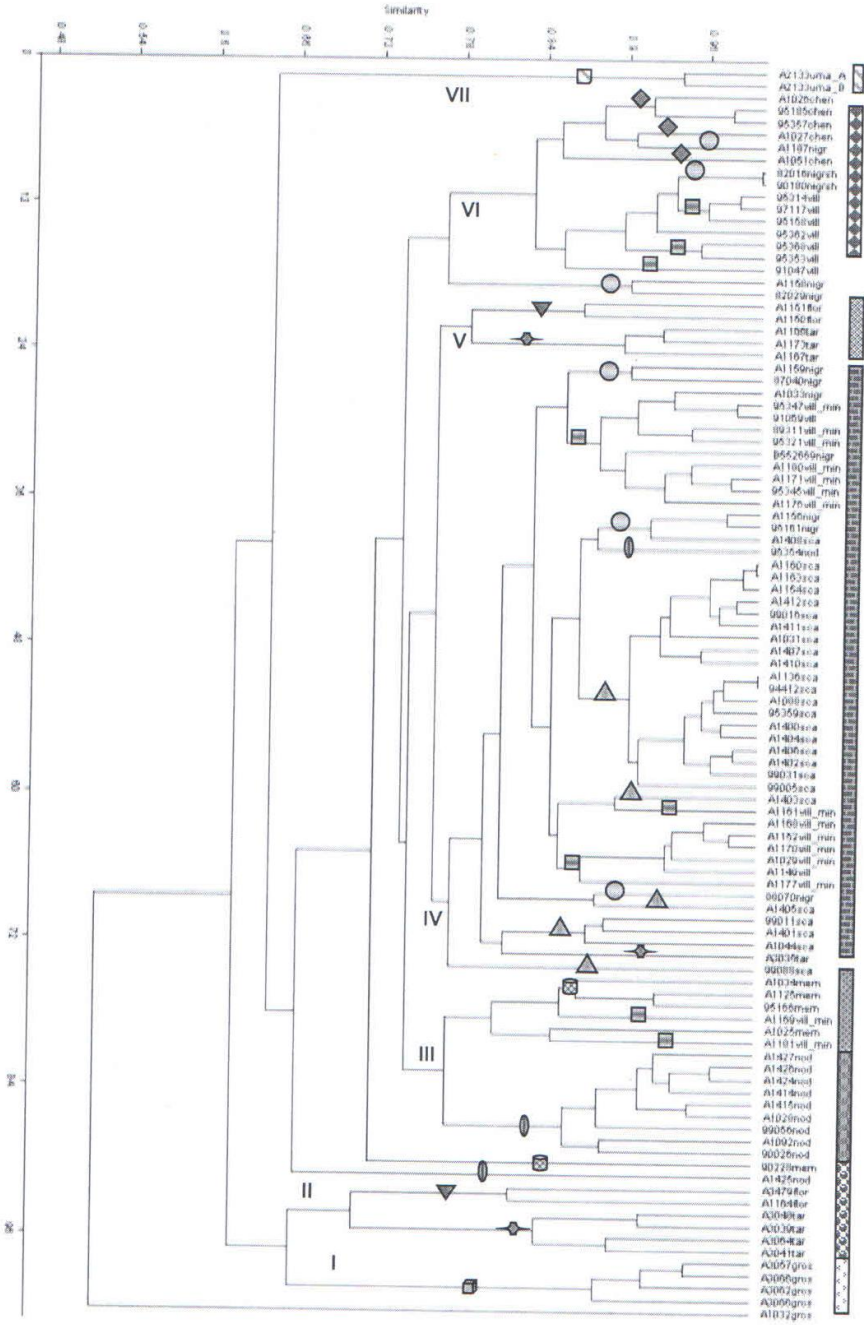


Figure 3: Dendrogram with the highest cophenetic coefficient produced based on a conventionally coded matrix analyzed with Rho coefficient. Signs on branches of each cluster indicate a mixture of individuals from different species in a single cluster**.**

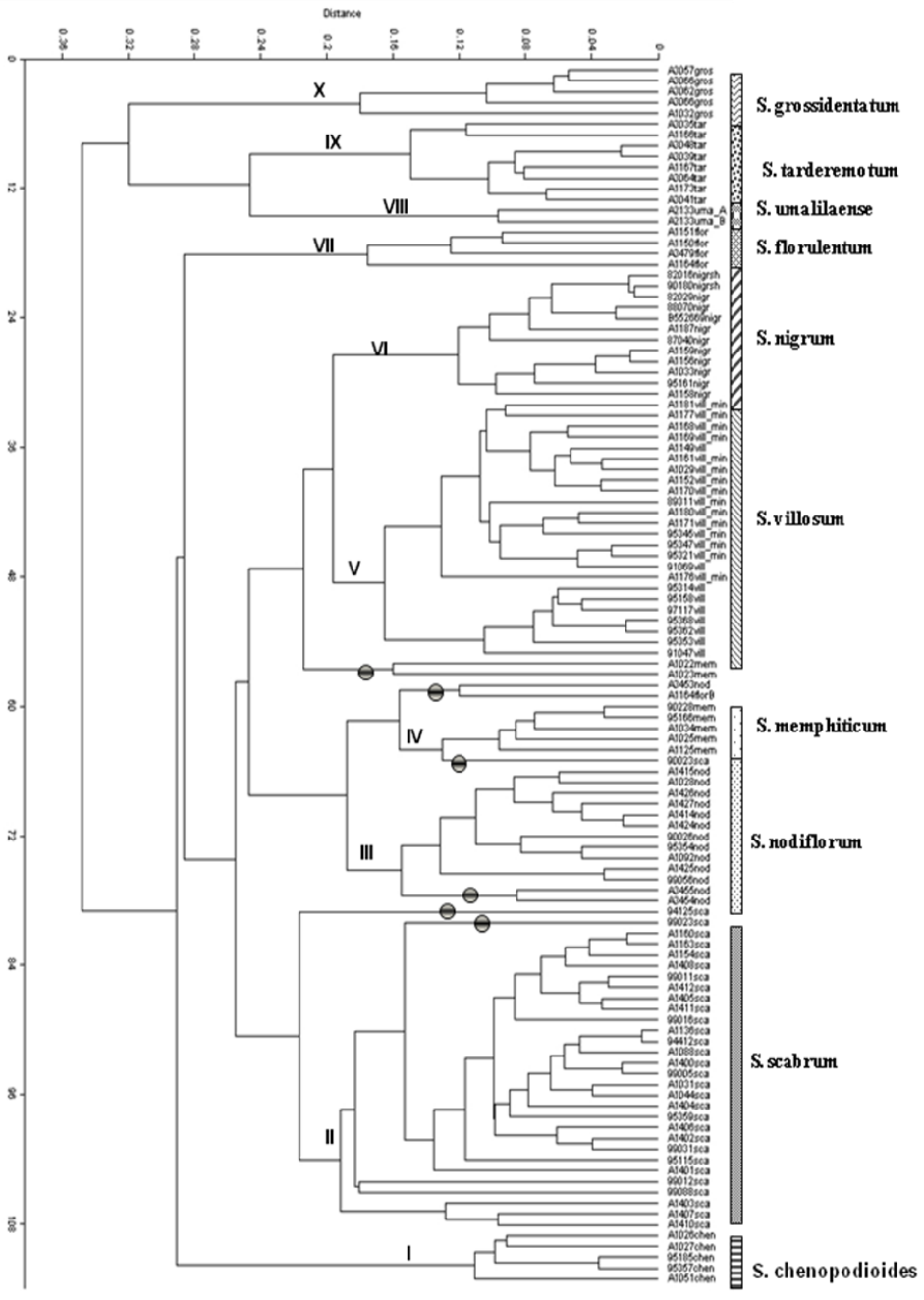


Figure 4: A dendrogram produced based on conventionally coded matrix including outliers analyzed using Gower's distance coefficient. The grey balls on the branches indicate position of identified outliers

DISCUSSION

Results in the present study shows that in phenetic analysis both coefficient of similarity and methods of coding of multistate characteristics affects clustering pattern. In fact, based on the current study similarity coefficients used can be divided into three groups. Group one is made up of Gower's and Hamming's coefficients that recovered all the ten species that were being tested regardless of the method of coding of data used. Group two is made up of Dice and Jaccard's coefficients these recovered the ten species being tested only when multistate characteristics were coded binary. The last group is made up of Manhattan, Rho, Euclidean coefficients and a dendrogram analyzed using mixed coefficients. These failed to recover any or recovered only a limited number of species even when the coding methods were changed.

This indicates that Hamming and Gower coefficients since they recovered all the ten species despite the change of coding method the two coefficients can be considered coefficients of choice under similar circumstances. Manhattan, Euclidean, Rho coefficients and the mixed coefficients option are probably not good choices to use when data are coded using binary or conventional methods. On the other hand, Jaccard's and Dice coefficients behave different from all other coefficients. The two coefficients produced results similar to Hamming and Gower coefficients when data were coded binary but failing like Euclidean, Manhattan and Rho with conventionally coded data. According to Jackson et al. (1989) Dice coefficient works like Jaccard's except that it gives more weight to the paired presence.

The above observations indicate therefore that selection of coefficients to use and method of coding should not be done blindly because they influence pattern of clustering and therefore the resulting classification. These results are in line with the conclusion made by Sokal (1986) and Jackson et al. (1989) though they did not test the coefficients that were tested in the present study. The pattern exhibited by different coefficients is probably expected. Different coefficients have been created for different purposes and are based on different algorithms. For example, although Gower, Euclidean, Manhattan are all distance coefficients, Gower is a general coefficient preferred with mixed data, it is thus the default measure for continuous and ordinal data. Manhattan is a geometric coefficient preferred when the interest is on individual characters. It is on this reason Manhattan is frequently used in phylogenetic inference with Wagner and Camin-Sokal procedures (Sokal 1986).

Based on the current study the difference between dendrograms produced by Gower and Manhattan is considerably larger. Whereas Gower recovered all ten species regardless of the coding method used, one of the two Manhattan's dendrograms recovered one species and the other zero. Sokal (1986) suspected that the differences of the resulting dendrograms produced using Manhattan and Gower coefficients was slight.

On the other hand, the poor performance of Euclidean coefficient one of the frequently used coefficient was predictable. The coefficient is not appropriate for data set with a mixture of data types e.g. continuous and nominal or ordinal (Finch 2005). A similar conclusion has also been reached with DNA markers (Kosman and Leonard 2005). The later authors showed that different similarity coefficients were also useful for different types of molecular markers.

Failure to form species specific clusters and thus not recovering the expected species is similarity of coefficient specific. For example, whereas Manhattan, Euclidean and Rho coefficients exhibited this habit, others that is Hamming and Gower exhibited the opposite. Thus though in some instances, failure to recover species has been attributed to lack of fit of classification to the data set based on the current study lack of fit may apply to some but not all dendrograms. Cophenetic correlation scores which measure the degree fit, shows that the dendrograms created based on conventional coded data set analyzed by Rho's coefficient scored the highest cophenetic value. Nevertheless, it is this dendrogram that had only two species specific clusters. Cophenetic correlation scores, have been used as a suitability index to select dendrograms that represents the classification better (for example, Conçalves et al. 2008). Based on the present study, this can only be true if the appropriate coefficient is used and in some instances if method of coding multistate data has been considered. Performance of Jaccard's and Dice coefficients evidences latter fact. Thus as Williams and Clifford (1971) and Holgersson (1978) suggests cophenetic correlation coefficient scores should not be taken without reservations.

Section *Solanum* has been considered one of the taxonomically complex groups for decades and reasons for the same have been given. However, based on the present study, the use of morphological data to delimit species in this group is practical. Mallet's (2007) argument that phenetic classifications based on morphology could group unrelated forms into paraphyletic or even polyphyletic taxa is disapproved. Lack of clustering of individuals of same species together in the present study was a function of coefficient of similarity and coding method. Other individuals that stayed away from your true identity were identified to be outliers (Figure 1). Actually, in the present study where the right similarity coefficient was use and/or with the correct method of coding and outliers removed, all individuals of same species regardless of their numbers clustered together first before they clustered with individuals of other species. This what is exhibited by Figure 2.

On the other hand, performance of UPGMA and of Gower coefficient in the present study was consistent with results obtained RAPDS markers obtained by Gonçalves et al. (2008).

To conclude, based on the present study it appears therefore that classification using phenetic approach is influenced mostly by coefficient of similarity and method of coding. The complexity of delimiting species as they have been observed in complex group such as section *Solanum* thus should not be attributed to morphology.

ACKNOWLEDGEMENT

The study was sponsored by ENVIRON program involving the University of Nijmegen (now Radboud Universality Nijmegen) and the Faculty of Science (now College of Natural and Applied Sciences) University of Dar es Salaam. Prof. Brita Stedje of Natural History Museum Norway gave me advices during the inception of the idea of writing this paper. Plants were grown at Radboud University Botanical garden.

REFERENCES

Bohs L and Olmstead RG 1997 Phylogenetic relationships in *Solanum*(Solanaceae) based on *ndh*F sequences. *Syst. Bot.* **22**:5-17.

Bohs L 2005 Major clades in *Solanum* based on *ndh*F sequence data. In: Keating, R.C., Hollowell VC and Croat TB, (eds.) A Festschrift for William G. D'Arcy, *The Legacy of a Taxonomist*. Missouri Botanical Garden, St. Louis.

Brower AVZ 2006 Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). *Syst. Biodvers.* **4***:*127–132.

Clement WL and Donoghue MJ 2012 Barcoding success as a function of phylogenetic relatedness in *Viburnum*, a clade of woody angiosperms. *BMC Evol. Biol.* 12:73doi:186/1471-2148-12-73.

Conçalves LS, Rodrigues R, do Amaral Jr. AT, Karasawa M and Sudre CP 2008 Comparison of multivariate statistical algorithms to cluster tomato heirloom accessions. *Genet Mol. Res.***7**:1289-1297.

Datwyler SL and Wolfe AD 2004 Phylogenetic Relationships and Morphological Evolution in *Penstemon* Subg. *Dasanthera* (Veronicaceae). *Syst. Bot.* **165**:165-176.

Edmonds JM 2012 Solanaceae In: Beentje HJ (ed.) *Flora of Tropical East Africa*. Royal Botanical Gardens, Kew.

Edmonds JM. 2007. Forsskål's *Solanum* section *Solanum* (Solanaceae) specimens. *Kew Bulletin* **62**:657- 670.

Edmonds JM 1978 Numerical Taxonomic studies of Solanum L. section Solanum (Maurella). *Bot. J. Linn. Soc*. **76**: 27-51.

Edmonds JM and Chweya JA 1997 Black nightshades. *Solanum nigrum* L. and related species. Institute of Plant Genetics and Crop Plant Research Gatersleben /International Plant Genetic Resources Institute, Rome, Italy.

Farris J 1969 On the cophenetic correlation coefficient. *Syst. Zool.* **18**:279-285.

Forey PL and Kitching IJ 2000Experiments in coding multistate characters. In: Scotland R, Pennington RT, (eds.) *Homology and Systematics. Coding characters for phylogenetic analysis.* London: Taylor & Francis.

Finch H 2005 Comparison of distance measures in cluster analysis with dichotomous data. *J Data Sci*. **3**:85-100.

Hebert PDN, Cywinska A, Shelley LB, Jeremy R and de Waard JR (2003) Biological identifications through DNA barcodes. *P Roy. Soc. Lond. B Bi*. **270**:313-321.

Hebert PDN, Penton EH, Burns JM, Janzen DH and Hallwachs W 2004 Ten species in one: DNA bar coding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator. Proc. Natl. Acad. Sci. U.S.A.* 101:14812-14817.

Jackson DA, Somer KM and Harvey HH 1986 Similarity coefficients: measure of co-occurrence and Association or simply measure of occurrence. *Am. Nat*. **133**:436-453.

Hammer Ø, Harper DAT and Ryan PD 2001 PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontol. Electron*. **4**: 1-9

Heo KS and Lim KT 2005 Glycoprotein isolated from *Solanum nigrum* L. modulates the apoptotic-related signals in 12-*0*-Tetradecanoylphorbol 13-Acetate-stimulated MCF-7 cells. *J. Med. Food*. **8**:69-77.

Holgersson M 1978 The limited value of cophenetic correlation as a clustering criterion. *Pattern Recogn.***10**:287-295.

Kamoun S, Huitema E and Vleeshouwers VGAA 1999 Resistance to Oomycetes: a general role of the hypersensitive response? *Trends Plant Sci.* **4**:196-200.

Knapp S 2001 Is morphology dead in the *Solanum* taxonomy? In: van den Berg RG, Barendse GWM, van der Weerden GM, Mariani C (eds.) *Solanaceae V*: Advances in Taxonomy and Utilization. University Press, Nijmegen.

Kosman E and Leonard KJ 2005 Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid and polyploid species. *Mol. Ecol.***14**:415-424

Levin RA, Watson K and Bohs L 2005 A four-gene study of evolutionary relationships in *Solanum* section *Acanthophora*. *Am. J. Bot*. **92**:603-612.

Levin RA, Myers NR and Bohs L 2006 Phylogenetic relationships among the "spiny Solanums" (*Solanum* subgenus *Leptostemonum*, Solanaceae). *Am. J. Bot*. **93**:157-169.

Liu J, Mollers M, Gao L, Zhang De-Qand Li De-Z 2010 DNA bar coding for the discrimintion of Eurassia yews (Taxus L, Taxaceaew) and discovery of Cryptic species. *Mol. Ecol. Resour.* **11**:89-100.

Mallet J 2007 Species concept of. In Levin S. et al. (eds.) *Encyclopedia of Biodiversity* **5**:427-440.

Manoko LKM 2007 *A systematic study of African Solanum L. section Solanum (Solanaceae)*. PhD. Thesis, The Netherlands.

Manoko LKM, van den Berg RG, Feron RMC, van der Weerden GM and Mariani C 2007AFLP markers support separation of *Solanum nodiflorum* from *Solanum americanum* sensu stricto (Solanaceae). *Plant Syst. Evol.***267**: 1-11.

Manoko LKM, van der Weerden GM, van den Berg RG and Mariani C 2012A new tetraploid species of *Solanum* sect. *Solanum* (Solanaceae) from Tanzania.*PhytoKeys***:16**: 65-74. DOI**:** 10.3897/phytokeys.16.2884.

Olet EA 2004 Taxonomy of *Solanum* L. Section *Solanum* in Uganda. PhD Thesis, Agricultural University of Norway.

Olet EA, Lye K and Heun M 2011 Amplified fragment length polymorphisms (AFLPs) analysis of species of *Solanum* section *Solanum* (*Solanaceae*) from Uganda. *Afr. J. Biotechnol*. **10:** 6387-6395.

Perez RMG, Perez JAL, Garcia LMD and Sossa HM 1998 Neuropharmacological activity of *Solanum nigrum* fruit. *J. Ethnopharmacol*. **62**:43–48.

Pons J, Barraclough TG, Gomez-Zurita J, Cardos A, Duran DP,Hazzel S. Kamoun S, Sumlin WD and Vogler AP 2006 Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. *Syst. Biol.* **55**:595–609.

Roddick JG 1991 The importance of the Solanaceae in medicine and drug therapy. In: Hawkes JG, Lester RN, Nee M and Estrada-R N (eds.) *Solanaceae III: Taxonomy, Chemistry, Evolution*. The Royal Botanical Gardens, Kew, U.K.

Sass C, Little DP, Stevenson DW and Specht CD 2007 DNA Barcoding in the Cycadales: Testing the Potential of Proposed Barcoding Markers for Species Identification of Cycads. *PLoS ONE 2(11): e1154.*

Simmons NB and Geisler JH 2002 Sensitivity analysis of different methods of coding taxonomic polymorphism: an example from higher-level bat phylogeny. *Cladistics* 18:571-584.

Singh SP, Raghavendra K, Singh RK and Subbarao SK 2001Studies on larvicidal properties of leaf extract of *Solanum nigrum* Linn. (family Solanaceae). *Curr. Sci*. **81**:1529-1530.

Sokal RR, Rohlf FJ and Kansas L 1962The comparison of dendrograms by objective methods. *Taxon* **XI**: 33 – 40.

Sokal RR 1986 Phenetic taxonomy: Theory and Methods. *Annu Rev. Ecol. Evol. Syst.* **17**:423-442.

Spooner DM, van den Berg, RG, Rodriguez A, Bamberg J, Hijmans RJ and Lara-Cabrera S 2004 Wild Potatoes (*Solanum* section *Petota*; Solanaceae) of North and Central America. *System. Bot. Monogr*. **68**:1-209.

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C and Willerslev E 2007 Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* 35 (3) e14doi:10.1093/nar/gkl938.

Weese TL and Bohs L 2007 A Three-Gene Phylogeny of the Genus Solanum (Solanaceae). *Syst. Bot*. **32**:445-463. 2007.

Wiens J 2000 Coding morphological variation within species and higher taxa

for Phylogenetic analysis In: Wiens J, (ed) *Phylogenetic Analysis of Morphological data*. Smithsonian Institution Press. Washington and London.

Williams WT, and HT and Clifford HT 1971 On the Comparison of Two Classifications of the Same Set of Elements. *Taxon*. **20**:519-522.

Xu J, Yin H and Li X (2009) Protective effects of proline against cadmium toxicity in micropropagated hyperaccumulator *Solanum nigrum*. *Plant Cell Rep*. **28**:325-333.

Zengfu X 2001 Proteinase inhibitor II from *Solanum americanum*: *molecular characterization and potential use in generating insect-resistance transgenic vegetables.* PhD Thesis, University of Hong Kong, Hong Kong.