Wendy J. Phillips-Rodriguez – Christopher J. Howe – Heather F. Windram

Some Considerations about Bifurcation in Diagrams Representing the Written Transmission of the Mahābhārata*

One of the basic characteristics of traditional stemmatology or text genealogy is its tendency to go from unity to multiplicity mostly by means of bifurcation. That means that a great majority of stemmata tend to appear as bifid trees with only two branches at every splitting event. This issue has been concerning textual scholars for several centuries, but it was probably the French textual critic Joseph Bédier who made the first conspicuous attempt at identifying it. In 1890 Bédier published an edition of the medieval French poem "Le lai de l'ombre" following what is now called "stemmatics" or "Lachmann's method".[1] However, when he revised his work some twenty years later, with the purpose of making a new edition, it occurred to him that most of the stemmata he had ever made or seen (not only for "Le lai de l'ombre" but for all other works) ended up separating manuscripts into two big families.[2] After a long study of the case, he came to the conclusion that (Bédier 1913: XXVI)

> [t]oute entreprise de classement, passée ou future, des manuscrits d'un texte a conduit ou conduira presque fatalement l'opérateur à les répartir en deux familles seulement.

In Sanskrit textual criticism this phenomenon can be observed, for instance, in the traditional division of Mahābhārata (MBh) manuscripts into a Northern and a Southern group, which was proposed by Sukthankar in his Prolegomena to the Critical Edition (1933: XXX).

---

* The authors would like to thank the National Council for Science and Technology, Mexico (CONACyT) and the Leverhulme Trust for their generous support to carry out this research.

[1] Strictly speaking, Karl Lachmann never proposed the method that takes its name after him (Salemans 2000: 19).

[2] The problem that many stemmata consist of consecutively bifurcated lines of transmission was addressed by Maas (1958: 47f.) and others. More recently the discussion was pushed forward by Timpanaro (1963, transl. into English in 2005: 157-187), who was, however, criticised by Reeve (1986). Timpanaro acknowledges that Reeve's critique is largely justified; cf. the chapter "Additional Materials A [Final Remarks on Bipartite Stemmas]" in Timpanaro 2005: 207-215.

Wiener Zeitschrift für die Kunde Südasiens / Vienna Journal of South Asian Studies, Bd. LII–LIII/2009-2010, 29-43

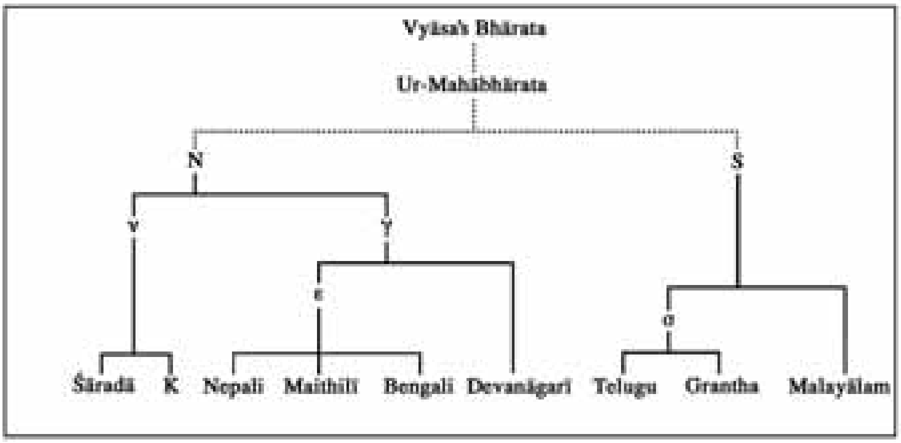

Figure 1: Pedigree for the Ādiparvan proposed by V. Sukthankar

As we can see, at the end of the branches of the diagram we find the groups of extant manuscripts collated for the Critical Edition (CE). For practical reasons, the editors of the CE first decided to separate the manuscripts into script groups (even though they were aware that regional features would only be preserved to a limited degree by manuscripts in the same script), and only after that did they proceed to the stemmatic analysis. Thus, the manuscripts were classified according to the script in which they were written, except for the group called K. The manuscripts of the K group are Devanāgarī manuscripts that preserve a distinct text, closely related to the Śāradā version. If we look further up the diagram we find Greek characters representing lost hyparchetypes (ν, γ, ε, and σ). The main identifiable level of branching is the big division of the manuscripts into a Northern (N) and a Southern (S) group. At the top of the tree we have the archetype from which all copies supposedly derived. Even though this is a pedigree and not a stemma proper,[3] it is very noticeable how every single branch splits always into two.[4]

---

[3] The difference between a stemma and a pedigree (e.g. the one proposed by Sukthankar) is the amount of information that each of them provides about the transmission. While a stemma is a diagram that shows the genealogical relationship of available and inferred witnesses, a pedigree only indicates the genealogical relationship of manuscript families.

[4] The only non-dichotomic division in the pedigree occurs at the level of the ε hyparchetype, which seems to be the ancestor of the Nepali, Maithilī and Bengali families. This division, however, was not made on stemmatic grounds but only with regard to the scripts. Therefore this apparent polytomy is not relevant at all to our discussion,

Historically, the division of the manuscripts into a Northern and a Southern group has been very influential. Even if Sukthankar proposed this division only for the Ādiparvan, it has long been considered both a characteristic of the MBh tradition and a true representation of its history of transmission. Nevertheless, not all the editors of the CE agreed with this division. Indeed, Franklin Edgerton, the editor of the Sabhāparvan, affirmed that (1944: XLVIII)

> it must be clearly understood that I do not regard "N" as a historic reality. I do not believe that [the] W[estern] and E[astern groups] are descended from a common secondary archetype.

In the last few years, several interdisciplinary projects have used techniques developed in the field of evolutionary biology to map the genealogical relationships between manuscripts (Barbrook et al. 1998, Howe et al. 2001). That is, they have been using algorithms based on the principles of cladistics and other evolutionary methods originally created for the purpose of mapping the relationships between organisms and inferring the evolutionary history of species from their DNA sequences. These are generally referred to as phylogenetic algorithms, and have achieved considerable success in the analysis of textual traditions. The basic statement of phylogenetics is that in a group of elements that come from the same ancestor there should be some shared primitive characteristics but also some derived characters. In this case, as in traditional stemmatics, the most revealing characters are the shared derived ones, the innovations, for they are responsible for the division or branching of the evolutionary diagram. In other words, what both evolutionary biology and textual criticism have in common is the principle that community of derived characters (we would rather say variants) reveals community of origin. As expected, the results of phylogenetic analyses occur in the shape of tree-like diagrams (or phylograms) where nodes imply ancestors and branches imply lines of descent. Phylogenetic methods are able to take into account every informative variant in each manuscript and can try millions of combinations and arrangements before hypothesising a number of possible stemmata. For example, a phylogram of some episodes of the Sabhāparvan looks like this:[5]

---

which is related to the classification of manuscripts according to their genealogical affiliations.

[5] The episodes analysed are *sarga*s 43-47, 51, 59, 60, 64 and 65, all of which belong to the Dyūtaparvan. The passage under investigation consists of 5235 characters (words or collatable units), of which 3249 characters are constant (i.e. no variant does occur). There are 1237 informative variants (characters which are present in at least two states,

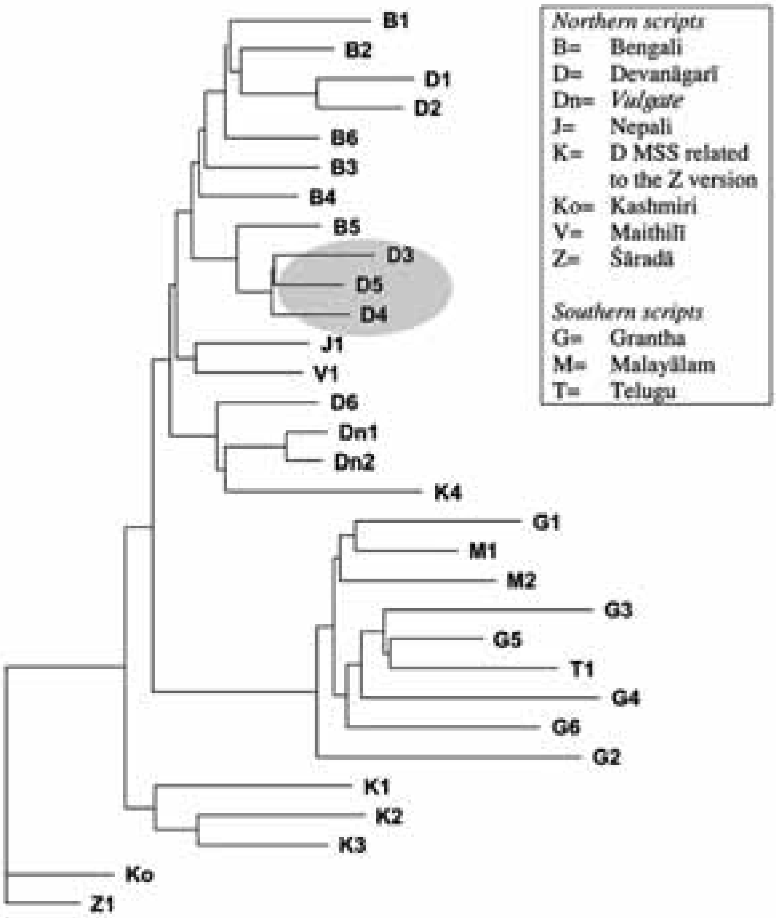

Figure 2: NJ phylogram[6]

We prepared this phylogram by feeding the program an accurate record of all the agreements and disagreements between the thirty-one manuscripts collated for the CE.[7] The tree illustrates several of the relationships previously hypothesised by the editors except that it describes

---

each of which is found in at least two manuscripts) and 749 non-informative variants (i.e. variants that concern only one single witness).

[6] All phylograms presented in this paper are unrooted; that means that no assumption about origin has been made. Thus the divergence separating Z1 and Ko from the other MSS does not necessarily imply they are closest to the archetype.

[7] Ko is a manuscript that Edgerton received only after he had completed his work on the critical apparatus of the Sabhāparvan. Since he found the textual quality of Ko remarkable, he added a full collation of the manuscript to his edition; see Appendix II in Edgerton 1944: 454-483.

them with further detail and infers a larger number of intermediary archetypes.

The most obvious differences between this diagram and the pedigree proposed in the first volume of the CE occur in the division of the so-called Eastern (E) family. First of all, K4 seems to be very close to Dn, which represents Nīlakaṇṭha's version. This is not a surprise since Sukthankar and Edgerton had already noticed that K4 is a hybrid manuscript that combines both Kashmirian and Vulgate features. Secondly, and more surprisingly, it can bee seen that Bengali and Devanāgarī manuscripts share such a large amount of features that the division into scripts, at least in their particular case, is not meaningful at all.

We would like to draw attention to the highlighted area, which shows a clear example of what we will call "consecutive bifurcation". We chose this particular area to show in a small scale what actually happens at all levels of the diagram.

There are several methods that may be used to build phylograms. For all diagrams we used a method called Neighbour-Joining (NJ) from the package PAUP* (Phylogenetic Analysis using Parsimony and Other Methods, Version 4). This algorithm uses a matrix of distances between witnesses that is constructed from the proportion of sites at which a pair of witnesses differ from each other. In other words, it proceeds by estimating the mean number of differences in specimens that have descended from a common ancestor, and then assigns a numeric value to the distance between each pair of them to make a pairwise distance matrix. Being an iterative algorithm, NJ works in a step-wise fashion. It starts from a starlike tree. Then, as its name indicates, it links the least-distant pair of nodes (neighbours) according to the distance matrix by creating a node that joins these two taxa. The next step is to calculate distances a) from each of the taxa in the pair to the new node and b) from all other taxa to the new node. Then the procedure starts again considering the pair of joined neighbours as a single taxon and using the distances calculated in the previous step. This is done recursively until all of the nodes are paired together and ancestral nodes are added accordingly.[8] The making of a distance matrix essentially is a data reduction from a many-state difference to a single number. However, NJ has proved its efficiency to work with large amounts of material that contain substantial differences. In general, it is able to reach results of equal and sometimes even greater accuracy than other methods for phylogenetic

---

[8] For a detailed step by step example of the procedure see Saitou – Nei 1987.

analysis. However, as Canettieri et al. (2005) point out, the problem is that since

> the methods are based on the comparison of sequences of characters and involve a grouping of those which appear closer to one another, it is obvious that in this way the trees are always dichotomic.

Thus, on account of their hierarchical structure, both the stemmatic model and the NJ method proceed by solving the first level of thinking (encouraging the division of the material into two groups only), and then move to the next level following exactly the same procedure. As a consequence, the resulting diagrams cannot allow for polytomies – that is, nodes that have more than two descendants (e.g. exemplars that may have more than only two copies). Trees built in this fashion could only deal with possible polytomies by means of consecutive bifurcations (as can be seen so clearly in the highlighted area in Figure 2). This, however, does not necessarily mean that those bifurcations are a reality of the textual tradition. As Bédier (1913: XXVIf.) has pointed out,

> [i]l serait merveilleux que le temps, en toute occasion et s'agissant d'une oeuvre littéraire quelconque ... se fût acharné à en détruire toute copie qui ne dérivait pas soit d'*x*, soit d'*y*, et que ce fût là une »loi« constante de l'histoire de la transmission des textes.

Nevertheless, we cannot blame the stemmatic model or the NJ method for their tendency to divide the material into dichotomies because that is exactly what they are asked to do. Actually, even though in many cases the dichotomies presented by the stemmata may not be a matter of reality but of methodological principle, this bifurcating tendency may be beneficial to the editorial process. Indeed, authors like Grier (1988: 272) have pointed out that

> the hazards of accepting a false multipartite stemma or the arbitrary elimination of a witness far outweigh those of retaining a false bipartite stemma.

In his 1988 paper Grier gives a clear explanation for this affirmation. If the editor chooses a multipartite stemma when the actual stemma of the textual tradition is bipartite, he is under risk of arbitrarily eliminating valuable witnesses or individual readings. On the other hand (1988: 277),

> even if the proposed bipartite relationships are false, they do not eliminate good readings, only those that would have been eliminated in any event by a true multipartite stemma, were it possible to demonstrate its existence, and potentially false readings are prevented from ascending to the archetype on stemmatic grounds.[9]

---

[9] For a very clear example see Grier 1988: 265.

Nevertheless, textual scholarship is not only about establishing a text but also about knowing more about its history of transmission. With the help of certain features of the new electronic tools there is a way to test the level of reliability of the proposed dichotomies. The following phylogram was built using exactly the same data that we used for Figure 2:

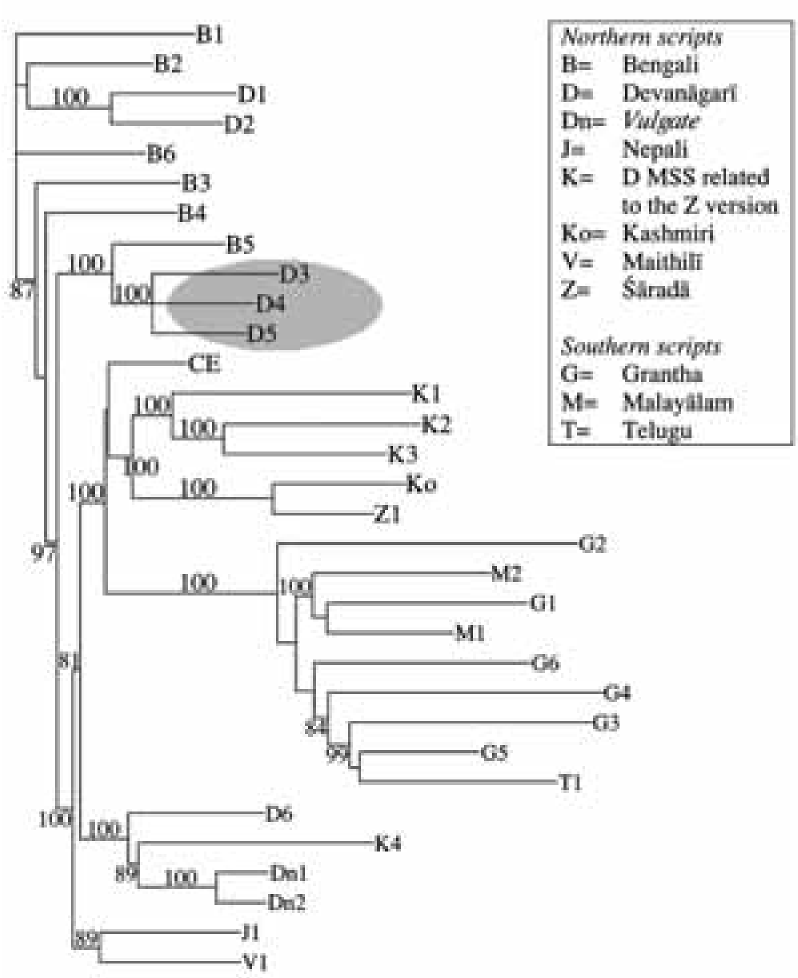

Figure 3: NJ bootstrap phylogram

Once again, the data was analysed using the NJ algorithm from the PAUP* package. In this case, however, we have also used "bootstrap analysis", a statistical procedure to test the reliability of the branching of the tree (we added the data derived from the text of the CE just to see how it compares with the data of the manuscripts). Bootstrapping is a resampling technique which uses a large number of replicate data sets, generated from the original data set by a random sampling method. The rationale behind this procedure is that since the reliability of a tree could only be tested in comparison with a different data set (which is impossible, because we only have this one single data set) the data has to be randomly resampled in order to get virtual alternative data sets. A phylogenetic tree is generated for each data set and the results are represented as a consensus tree which only shows manuscript groupings supported by at least 50% of the individual trees (in Figure 3 only bootstrap support values of more than 80% are labelled). By resampling the material 1000 times we make sure that we filter out most hypotheses built on poorly supported affiliations and retain only those groupings that have a more robust support. Bootstrap analysis thus allows for the estimation of confidence levels for particular groups.

As can be seen, one of the most interesting features of trees built with bootstrap analysis is that they can produce images where weakly supported dichotomies appear as polytomies. For example, in this case we can see a clear polytomy in the case of D3, D4 and D5. If we go back to Figure 2 we can see how it is dealt with by proposing two bifurcating divisions instead of just one polytomy with three branches. What has happened in Figure 2 is that the number of shared variants between D3 and D5 is slightly larger than the number of variants each of them shares with D4. As a consequence they have been assimilated into a single family. However, the bootstrap analysis has been able to identify the fact that such consecutive dichotomies are only weakly supported. It must be clear that the bootstrap resampling technique is not an alternative phylogenetic method in its own right, and does not address any weaknesses in the assumptions of the method. Thus, if NJ (or any other phylogenetic method) has a methodological constraint in tree generation (e.g. producing bifurcating trees), then bootstrapping will not cure this. All that bootstrapping does is to highlight the fact that some of the bifurcations fall beneath our chosen support threshold, so they are not shown in the bootstrap tree. Thus, it is important to keep in mind that the sort of polytomies presented by bootstrap analysis do not establish a multipartite stemma. They only question the certainty of a dichotomy. In a way, polytomies of this sort just bring more uncertainty to our stemmata. By replacing weak dichotomies, polytomies invite the scholar to be even more cautious about what to expect from a *stemma codicum*.

If we take this sort of example to a further level (from single manuscripts to families of manuscripts), we will realize that the double bifurcation that seems to lie right at the core of Sukthankar's division of the MBh manuscripts (Ur-MBh into N and S, then N into E and W) could be in need of revision with the help of the new tools.

One of the most interesting possibilities offered by phylogenetic tools is that they allow various representations of the results. To give a more graphic idea of what we mean let us show exactly the same results of Figure 3 displayed in a different manner:

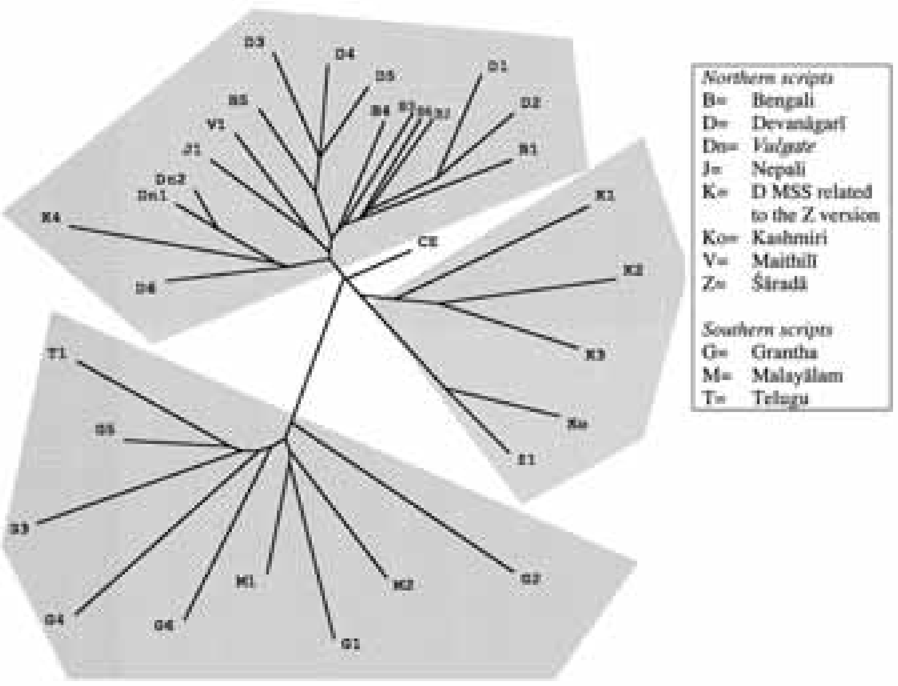

Figure 4: NJ bootstrap phylogram (alternative representation)

The division of manuscript families is identical with the one shown in Figure 3 (where the same tree appears more as a vertical ladder). Shown like this, however, it is easier to see that the data support a division into three main branches each one corresponding to a different manuscript family: Eastern, Western and Southern. Certainly the Eastern and the Western families are closer to each other in terms of branch length than either of them is to the Southern. However, that does not necessarily mean that they both (E and W) share one single hyparchetype.

The division of the MBh transmission into a Northern and a Southern recension could be the result of a certain (not completely accurate) way of interpreting the information. To simplify, we could say that Figure 4 shows a diagram of the following sort (Figure 5 where E represents the Eastern family of manuscripts, W the Western family and S the Southern family):

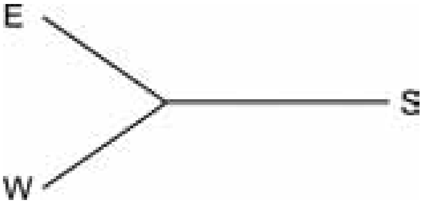

Figure 5: Simplified unrooted diagram

As can be seen we have three leaves (E, W and S). E and W are closer to each other than either of them is to S. That, however, does not necessarily mean that they both belong to a family of which S is not part. Indeed, the diagram does not suggest at all that E and W share a hyparchetype. It only states that the differences between them are fewer than between any of them and S.

However, things seem to be different if we try to add a vertex to represent the archetype ("root"). There is a tendency to believe that once we add a root to a diagram like the one in Figure 5 the resulting figure would be of the following sort:

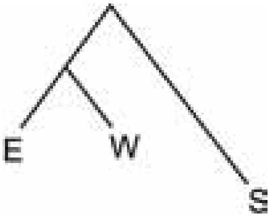

Figure 6: Attempt to root the diagram presented in Figure 5

It seems that the very principle of "agreement in error", which lies at the basis of both traditional stemmatics and phylogenetics, can lead to this particular reasoning. In other words, if two groups of manuscripts share between them more conjunctive errors than either does with a third group the principle of "agreement in error" would grant them a common hyparchetype. Indeed, this line of thought could have given place to the division of the MBh into a Northern and a Southern family, with the Northern divided into an Eastern and a Western group. Nevertheless, this bifurcated division may not necessarily represent the true genealogy. It could be that two groups that are not a genuine family (E and W) have been traditionally grouped together just because they are very different from the third one (S). Indeed E and W share more common variants with each other than either of them does with the Southern family, which contains a significant amount of extra material not found in the other two groups. As a consequence, E and W could have been classified together as if they were a genuine family, the so-called Northern family. Nevertheless, it may be worth asking if in a complex tradition like that of the MBh, where contamination played a very important role, common variants are unquestionable indicators of genealogical relationships.[10] The fact that some families share more variants with one certain family than with others may sometimes occur by chance (parallelism), sometimes by contamination or sometimes because they are actually closer versions, but that does not necessarily mean that they share a common immediate written hyparchetype. Stemming from geographically close regions (that were connected by well established trade routes), the chances that contamination could have happened between manuscripts of the E and the W group were much higher than for contamination with manuscripts from the faraway South. The amount of shared contaminatory readings between manuscripts from the E and from the W group could be an important reason why those manuscripts traditionally have been considered to belong to a single family. Indeed, as Grier (1988: 268) mentions, following an argument by Timpanaro (1963),

> [a]ctual multipartite stemmata might appear to have fewer branches if significative errors have been removed from one or more witnesses through contamination.[11]

---

[10] It is a well known fact that manuscripts of the MBh were constantly revised against each other. This is evident by the great number of annotated manuscripts that are available. In fact, as Dunham (1991: 3) points out, "in quite a few instances it is possible to see how marginal glosses and variant readings were absorbed into the text itself through the practice of comparison". For specific examples see Dunham 1991.

[11] Reeve (1986: 66) makes a very interesting point about this same matter: "Contamination, it seems to me, is a process more likely to reduce the number of shared errors in a tradition than to increase it and therefore more likely to increase the apparent number of branches in a stemma than to reduce it." This may be true for some traditions. However, in the particular case of the MBh the long and very widespread practice of comparing and annotating manuscripts resulted in a certain degree of homogeneity between manuscripts of neighbouring areas, irrespective of their vertical stemmatic affiliations.

In fact, when we try to root the diagram in Figure 5, different options than the traditional bifurcation of MBh manuscripts into a Northern and a Southern family are possible and plausible. As mentioned before, nothing in our unrooted diagrams points at a particular rooting such as the one shown in Figure 6. Actually, one of the main characteristics of an unrooted tree is that, in the absence of additional information, its root could actually be anywhere in the tree. It does not necessarily have to be in the centre, or near the centre, but it could be anywhere along the branches or at the endpoints. Just to give an illustration of the many possibilities of rooting options of the diagram in Figure 5, we can show how the trees would look if the root were to be placed on any of the edges or at the internal node of the tree:

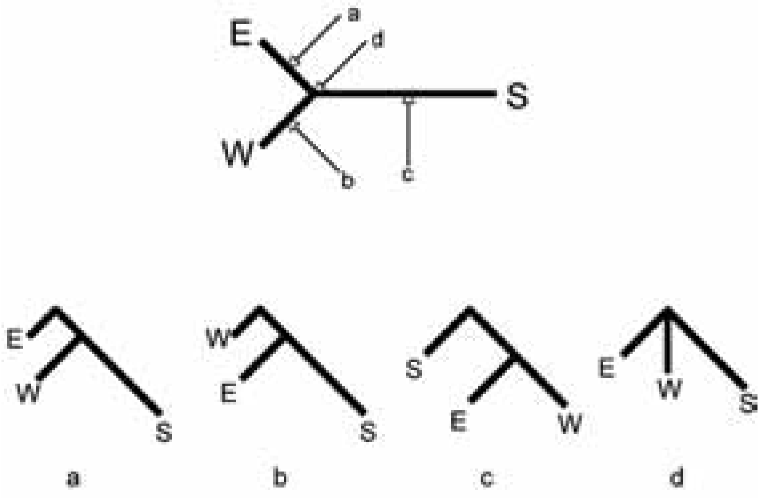

Figure 7: Four different rooting options

It is not possible to determine which of the possible diagrams represents the true history of transmission of the chapters analysed.[12] As far as the diagrams (Figures 4 and 5) are concerned, there is no particular reason to favour the interpretation of a bifurcation of the MBh tradition into a Northern and a Southern family over all other possibilities. This is a clear reminder that we cannot possibly arrive at a *stemma codicum* without a philological discussion of variant readings. No method, be it stemmatics or phylogenetics, can produce true stemmata by itself, with-

---

[12] Different parts of the MBh may have different histories of transmission.

out a careful exploration of the variants by the scholar. Edgerton, who is the one who can claim to have made the most detailed analysis of variant readings of the Sabhāparvan, does tell us that he believes to have found three streams of MBh tradition (1944: XL):

> W (a Western type), E (Eastern), and S (Southern). W includes Ś and K; E includes Ñ, V, B, and D; and S includes T, G, and M.

However, he worked only with the Sabhāparvan.[13] We are thus left with the task of finding out by ourselves, making use of all the tools we may have at hand, which of all possible diagrams represents best the actual history of the whole transmission.

It may seem that we have left the issue unfinished without answering the question of how the written transmission of the MBh split below the archetype, namely, into a bifurcation or into polytomy, or indeed whether a common archetype of all the versions available today did ever exist. For the moment, all we can do is to point out that the time-honoured ideas about the transmission of the MBh should not be taken for granted without further revision and to provide a clear example why such revision is needed. By investigating where our perception of this transmission comes from and questioning how accurate it is – not completely accurate as regards to bifurcation we may say –, we are laying the grounds for a finer-grained discussion of the subject. We trust that little by little the study of the available manuscripts will be able to unveil more information about the way the transmission unfolded.[14]

It is true that what we call "stemmatics" was first developed with the particular purpose of finding some guidance to establish a text. However, through time stemmata have become more than just guiding tools to choose readings. They also provide us with the chance to glance at the historical development of a text. The discussion about the different plausible scenarios that could have given rise to the division of the MBh manuscripts may not have been of any practical use to the editors who were establishing the text of the CE. However, as textual scholars we do care about such discussion. We care because it helps us understand better the process of transmission of our text and it gives us valuable clues about the way that manuscripts were disseminated in the Indian

---

[13] He also suggested the same state of affairs for the Ādiparvan, but did not deal with that matter at length.

[14] It should not go without mention that doubts concerning the reliability of Sukthankar's stemmatical hypothesis were already voiced by Grünendahl (1993), according to whom the Nepalese manuscript tradition deserves more attention than it received so far.

subcontinent. As Reeve (1986: 65) rightly points out, in textual criticism "methodology and history cannot of course be neatly separated" because both of them are essential to the study of our text. For the sake of methodology we may accept bipartite stemmata; however, for the sake of history we must try to test to what extent those bifurcations are a reality of the textual tradition.

Bibliography and Abbreviations

Barbrook et al. 1998	A.C. Barbrook – N. Blake – P. Robinson, The Phylogeny of the Canterbury Tales. *Nature* 394 (1998) 839.
Bédier 1913	J. Bédier, Introduction. In: J. Renart, *Le lai de l'ombre*. Paris: Librarie de Firmin-Didot et Cie, 1913.
Canettieri et al. 2005	P. Canettieri – V. Loreto – M. Rovetta – G. Santini, Higher Criticism and Information Theory. *Rivista di Filologia Cognitiva,* http://w3.uniroma1.it/cogfil/ecdotica.html (2005).
CE	Critical Edition of the Mahābhārata.
Dunham 1991	J. Dunham, Manuscripts Used in the Critical Edition of the Mahabharata. A Survey and Discussion. In: A. Sharma (ed.), *Essays on the Mahabharata*. Leiden: E.J. Brill, 1991, p. 1-18.
Edgerton 1944	F. Edgerton (ed.), *Sabhāparvan. The Mahābhārata* (Vol. 2). Poona: Bhandarkar Oriental Research Institute, 1944.
Grier 1988	J. Grier, Lachmann, Bédier and the Bipartite Stemma: Towards a Responsible Application of the Common-Error Method. *Revue d'histoire des textes* 18 (1988) 263-278.
Grünendahl 1993	R. Grünendahl, Zur Klassifizierung von Mahābhārata-Handschriften. In: R. Grünendahl – J.-U. Hartmann – P. Kieffer-Pülz (ed.), *Studien zur Indologie und Buddhismuskunde*. Festgabe des Seminars für Indologie und Buddhismuskunde für Professor Dr. Heinz Bechert zum 60. Geburtstag am 26. Juni 1992. [*Indica et Tibetica* 22]. Bonn: Indica et Tibetica Verlag, 1993, p. 101-130.
Howe et al. 2001	C.J. Howe – A.C. Barbrook – M. Spencer – P. Robinson – B. Bordalejo – L.R. Mooney, Manuscript Evolution. *Trends in Genetics* 17 (2001) 147-152.
Maas 1958	P. Maas, *Textual Criticism*. Transl. from the German by B. Flower. Oxford: Clarendon, 1958.
MBh	Mahābhārata: *The Mahābhārata*. For the First Time Critically ed. by V.S. Sukthankar, S.K. Belvalkar et al. 20 vols. Poona: Bhandarkar Oriental Research Institute, 1933(1927)-1966.
NJ	Neighbour-Joining. Cf. Saitou – Nei 1987.
PAUP*	D.L. Swofford, *PAUP**. Phylogenetic Analysis Using Parsimony (*and other methods). Sunderland, MA: Sinauer Associates, 2001.

Reeve 1986	M.D. Reeve, Stemmatic Method: "Qualcosa che non funziona"? In: P. Ganz (ed.), *The Role of the Book in Medieval Culture*. Proceedings of the Oxford International Symposium, 26 September – 1 October 1982. Brepols: Turnhout, 1986, I/57-70.
Saitou – Nei 1987	N. Saitou – M. Nei, Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution* 4,4 (1987) 106-425.
Salemans 2000	B.J.P. Salemans, *Building Stemmas with the Computer in a Cladistic, Neo-Lachmannian Way*. Nijmegen: Katholieke Universiteit, 2000.
Sukthankar 1933	V. Sukthankar (ed.), *Ādiparvan*. *The Mahābhārata* (Vol. 1). Poona: Bhandarkar Oriental Research Institute, 1933.
Timpanaro 1963	S. Timpanaro, Stemmi bipartiti e contaminazione. In: Id., *La genesi del metodo del Lachmann*. Firenze: Felice le Monnier, 1963, p. 112-135.
Timpanaro 2005	Id., *The Genesis of Lachmann's Method*. Ed. and transl. by G.W. Most. Chicago – London: University of Chicago, 2005.