



**HAL**  
open science

## Polyphemus: a Palaeolithic Tale?

Julien d'Huy

► **To cite this version:**

Julien d'Huy. Polyphemus: a Palaeolithic Tale?. The Retrospective Methods Network Newsletter, 2015, 9, pp.43-64. halshs-01170837

**HAL Id: halshs-01170837**

**<https://shs.hal.science/halshs-01170837>**

Submitted on 6 Jul 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

- Heusler, Andreas, & Wilhelm Ranisch (ed.). 1903. *Eddica Minora*. Dortmund: Druck und Verlag von Fr. Wilh. Ruhfus.
- Jón Helgason. 1953. "Norges og Islands digtning". In *Litteraturhistorie B: Norge og Island*. Ed. Sigurður Nordal. Nordisk kultur 8. Stockholm: Albert Bonniers Förlag. Pp. 3–179.
- Kock, Ernst Albin. 1923–1944. *Notationes Norrænæ: Anteckningar till Edda och skaldediktning I–XXVIII*. Lunds Universitets årsskrift new ser. 1. Lund: Gleerup.
- Leslie, Helen F. 2012 [2013]. *The Prose Contexts of Eddic Poetry: Primarily in the Fornaldarsögur*. Dissertation for the degree of philosophiae doctor (PhD). Norway: University of Bergen.
- Lexicon Poeticum: Antiquæ Linguae Septentrionalis – Ordbog over det norsk-islandske skjaldesprog*. 1931. Ed. Sveinbjörn Egilsson & Finnur Jónsson. Copenhagen: Møllers.
- Magnus Magnusson & Hermann Pálsson (trans.). 1966. *King Harald's Saga: Harald Hardradi of Norway: From Snorri Sturluson's Heimskringla*. Harmondsworth: Penguin.
- Nordal, Sigurður, et al. (eds.). 1944. *Flateyjarbók I*. Akraness: Flateyjarútgáfan.
- Rowe, Elizabeth Ashman. 2002. "Sǫrla þáttir: The Literary Adaptation of Myth and Legend". *Saga-Book* 26: 38–66.
- Stavnem, Rolf (ed.). 2012. "Rekstefja". In *Poetry from the Kings' Sagas I:2*. Ed. Diana Whaley. Skaldic Poetry of the Scandinavian Middle Ages 1. Turnhout: Brepols. Pp. 893–939.
- Sveinbjörn Egilsson (ed.). 1825–1837. "Haralds saga Harðráða". In *Fornmanna sögur eptir gömlum handritum útgefnar að tilhlutun hins norræna fornfræða félags VI*. Copenhagen: Popp.
- Townend, Matthew (ed.). 2009. "Haraldsstikki". In *In Gade* 2009b: 807–808.

## Polyphemus: A Palaeolithic Tale?

Julien d'Huy, Institute of the African World (IMAF), Paris I Sorbonne

*Abstract: This paper presents an analysis of 56 variants of European and North American examples of the so-called Polyphemus tale (international tale type ATU 1137) using phylogenetic software according to 190 traits. Discussion addresses a number of points of comparative methodology while considering the historical implications of a relationship between different versions of this tale type recorded in diverse cultures.*

Les objets qui posent à l'ethnologue un problème de classification sont certes moins nombreux que ceux soumis à l'attention des naturalistes. L'ethnologue n'en a que plus de raisons de chercher des enseignements peut-être, des stimulations certainement, auprès de disciplines qui travaillent sur les mêmes problèmes à une échelle incomparablement plus grande et avec des méthodes plus rigoureuses. (Lévi-Strauss 2002: 311.)

The objects that pose a problem of classification to the ethnologist are certainly less numerous than those brought to the attention of the naturalists. The ethnologist also has all the more reasons to look perhaps for lessons, certainly for stimulation, from disciplines that work on the same problems on an incomparably larger scale and with more rigorous methods. (My translation.)

The Finnish School of comparative folklore research has an empirical and positivistic approach to using the so-called Historical-Geographic Method and its variations, which was recently discussed by Frog in an earlier volume of this journal (2013b). The scholars of this school tried to collect all variants of a tale, to analyse the diffusion and frequency of each of its individual traits, and to trace each

motif's history and geographical spread. They also tried to reconstruct the ideal primeval form of the tale (*Urmärchen*) from which all the attested versions ultimately originated. Despite an initial enthusiasm, the reconstructive ambitions of the Finnish School have been strongly criticized. This method was conceptualized long before the development of computer-assisted methods, which may hold some potential for revitalizing this type of research. The present article considers the potential use and value of applying modern phylogenetic tools for the study of myths and folktales.

### *The Biological Model*

A great advance in biology occurred when researchers realized that the lineage of organisms could be represented with a branching diagram or 'tree'. This structure visualizes the inferred evolutionary relationships among various biological species based upon similarities and differences in their physical or genetic characteristics. Each node from which branches of the tree stem represents a speciation event in which a lineage splits into two or more descendant lineages (i.e.

Table 1. Equivalence of elements and features in the comparison of genetic systems and of myths / folktales.

<b>Genetic Systems</b>	<b>Myths / Folktales</b>
Discrete heritable units (e.g. the four nucleotides, codons, genes and individual phenotypes)	Discrete heritable units (e.g. mythemes, motifs, tale-types)
Mechanisms of replication by transcription and reproduction	Teaching, learning and imitation
Slow rate of evolution	Fast or slow rate of evolution
Parent–offspring, occasionally clonal	Parent–offspring, intergenerational transmission, teaching, writing (more recent)
Mutation (e.g. slippage, point mutation and mobile DNA)	Innovation (e.g. variation, innovation, mistakes)
Natural selection of traits (individuals with certain variants of the trait may survive and reproduce more than individuals with other variants)	Social selection of traits (e.g. societal trends and conformist traditions)
Allopatric or sympatric speciation	Geographical or social separation
Hybridization	Mixture of two or more myths or tales
Horizontal transmission defined to be the movement of genetic material between bacteria or within the genome other than by descent in which information travels through the generations as the cell divides (e.g. viruses, transposons)	Extralineal borrowing or imposition
Geographic cline	Mythological transformations
Fossils	Ancient texts
Extinction	Disappearance

branches). Biological and mythological entities have many traits in common, as summarized in Table 1. The most important of these is the fact that both are formed by discrete heritable units which evolve progressively with time. The more two related species or two myths diverge, geographically and temporally, the more distant their genetic relationship probably is. Observing these parallels, software developed for assessing genetic relationships and relatedness can potentially be applied to assess corresponding relationships between examples of myths and folktales.

Applying the biological model to myths and tales is not new. According to Carl Wilhelm von Sydow (who himself was following a long tradition: see Hafstein 2005), folktales are like biological beings (von Sydow 1927; 1948 [1965]: 238–239): they tend to adapt to their environment and they evolve by means of natural selection. This explains why so many individual variants of tales differ from the abstract tale-types with which they are identified. As early as 1909, Arnold van Gennep stated that folkloric elements should be studied “comparatively,

with the aid of the biological method” (van Gennep 1909: 84).

Initial published attempts using phylogenetic software to study mythology and folktales may date back to 2001. Jun’ichi Oda applied an alignment program used for genome informatics to Propp’s “sequence of functions” concerning 45 fairy tales. Propp’s sequence was reduced using the Greimas model to 16 functions, each of which was in turn coded as a given amino acid (e.g. fairy tale 1 = ACDEF; fairy tale 2 = ACFDEF; fairy tale 3 = ADPHW, etc.). The use of a program could then arrange the sequences of functions to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences in the same way it worked for amino acids in a genome. This approach presents some difficulty owing to limits of the genome model: the researcher can only work with only a limited number of functions (only 20 types of amino acids exist) and results were limited to only very short lines of code for each sequence (from 4 to 12 functions / amino acids). Under such circumstances, it is highly possible that convergent evolution could produce apparent

similarity between functions that are evolutionarily unrelated. More research would be necessary to arrive at any certainty that is impossible to obtain with this method. An additional problem is that the results depend on the specific order of sequences, in which case a variation in a conventional plot whereby e.g. the function of a donor occurs early could make tales appear to correspond in their formal sequence of functions that otherwise have nothing to do with one another. Oda's work holds a position in the history of research, but the effectiveness of Oda's method was never tested, for instance by changing the dataset or the method in order to control the results.

As far as I know, I was the first, in the beginning of 2012, to tackle many of the remaining problems with this sort of approach (d'Huy 2012a–c; 2013a–e). I studied many families of mythological narratives and folktales using different datasets of mythemes each time (vs. Oda's functions; see the definition below). I used as large a sample of versions as possible and multiplied the most up-to-date statistical and phylogenetical methods applied. This work has been then continued by other researchers, such as Jahmshid Tehrani (2013) and Robert Ross, Simon J. Greenhill and Quentin D. Atkinson (2013). Phylogenetic methods have been used to study many folktales and myths, including Pygmalion, the Cosmic Hunt, Polyphemus, the Dragon, Little Red Riding Hood and the Kind and the Unkind Girl.<sup>1</sup> Indeed, the phylogenetic approach is very interesting. It can offer answers to a lot of questions. At its most basic, it can be used to explore the extent to which examples of a given folktale exhibit a tree-like set of relations, and this can be interpreted as reflecting the relative contributions of vertical and horizontal processes in folktale evolution (d'Huy 2012a–b; 2013a–c; 2013e–f; Ross et al. 2013). It can be questioned whether the members of a so-called tale-type or motif indeed form a unity or should better be regarded as divided into phylogenetically distinct international types (d'Huy 2013e; Tehrani 2013), and whether we can reconstruct the proto-tale and its evolution (d'Huy 2012b; 2013a–c; 2013e–f; 2014a; Tehrani 2013). When the tree-like

relations of variants of a tale or tales in a database of mythology appear interpretable as reflecting its historical spread through the world, it becomes possible to consider whether this correlates with reconstructions of human migrations that might be responsible for that spread (d'Huy 2012a–c; 2013a–c; 2013e–f; Tehrani 2013; d'Huy & Dupanloup 2015). More generally, the phylogenetic approach offers new resources for considering how folktales evolve (d'Huy 2013a; 2013c; 2013d; Ross et al. 2013). As the approaches of the Finnish School fell out of favour in the latter half of the 20<sup>th</sup> century and research paradigms changed, folkloristic research on folktales and myths moved away from questions about the history of tales and the historical relationships behind their various forms (Frog 2013b: 21–22). Returning to these questions now with the support of modern phylogenetic tools has the potential to produce new knowledge.

### ***Confronting Methodological Problems***

The Historical-Geographic Method (HGM), especially as it became internationally known in the first half of the 20<sup>th</sup> century or the 'Classic HGM' (esp. Krohn 1926), suffered from a number of methodological problems for which it received heavy criticism (Frog 2013b). Phylogenetic tools have the potential to resolve a lot of the problems addressed by its critics (d'Huy 2013a; Tehrani 2013). Several of these issues will be briefly reviewed here:

#### *1. It is impossible to reconstruct the tale as it was first composed and told to others.*

Phylogenetic tools statistically assess degrees of formal relatedness between items. Rather than shared mutations, the degrees of formal relatedness are hierarchically organized in a tree according to variations that they hold in common, which may be produced by historically shared innovations. This makes it possible to model the evolution of a tale inside a tree statistically. This approach is similar to the formal studies of the Classic HGM, but uses a computer rather than graph paper. It does not involve qualitative assessment of the features of variants and thus the statistical reconstruction is essentially a mathematical outcome of the correlation of

similarity of individual elements. Insofar as this method makes this statistical assessment quantitatively on the basis of the number of individual elements without being hierarchically structured according to larger units of narrative, it is (hypothetically) possible that variants could be grouped together owing to a concentration of formal similarity in the co-occurrence of motif elements in one episode even though the overall narrative form and structure was close to that of another set of variants. For this reason, the elements chosen for each motif need to be shared equitably throughout the whole story. Where formal relatedness of one example does not align with other shared variations of a group, the software makes this observable as a conflict in the data.

*2. The Classic HGM could not show how two or more seemingly different themes could stand in a structural transformational relationship to each other (Lévi-Strauss 1968: 185).*

The Classic HGM's focus on the presence or absence of story details neglected the logical relationships—evident—between different versions of a same myth. At least two additional principles (variation and selection) in the process of folklore transmission are compatible with both evolutionist and structural treatments: the more two myths diverge from each other or transform, the more distant their genetic relationship. This formal distance seems normally to correlate with geographical and/or temporal distance of the examples (e.g. Ross et al. 2013). However, phylogenetic tools allow for the process of divergence to occur more quickly in one region and more slowly in another. The use of phylogenetic tools also can take into account the fact that the tradition in one cultural area can undergo an abrupt and radical transformation that rapidly becomes socially dominant (e.g. with religious change). These tools assess formal relationships: the interpretation of the history behind that formal relatedness is a subsequent analysis by the researcher.

*3. The reconstructive approach identified variation with dispersal and reconstructing the historical form of a tale was thus linked to identifying its location of origin.*

Any attempt to find the place of origin of tales seems to be doomed to failure. The evidence of individual tales has not been evenly collected among all cultures and the narrative has the potential to be transmitted across different areas, carried via contact networks and population mobility. This process of transmission has the potential for even the repeated displacement of earlier local and cultural forms as a historical process. The tale may also simply drop out of use in some areas without leaving evidence of the local form, and there may not be any evidence to link a tale to the geographical area of its origin. Moreover, the geographical emphasis developed from “confusing a continuum of typological similarities [in the distribution of variants] with a historical progression of developments accompanying geographic spread” (Frog 2013a: 117), which is roughly like interpreting variation across dialects of a language as reflecting a sequence of developments based on the language's progressive spread to new locations. Such continua may be better understood as related to contact networks in interaction, moderating and negotiating variation. This phenomenon of cultural adjacence (Frog 2011: 92–93) could make tracing locations of origin and processes of geographical spread problematic and most often impossible without support from other types of evidence or association with a broader system of material (e.g. a cultural mythology). Phylogenetic tools can easily accommodate incomplete phylogeny. Moreover, some tools (such as midpoint rooting) may enable the essential features of a tale from which all of the variants derive to be established. However, phylogenetic tools treat formal relationships between texts and not their geographical distribution. A researcher may take the information produced in a phylogenetic analysis and compare that with the geography of formal distribution and the history of cultures from which examples were collected, but that is another level of analysis and interpretation.

#### 4. *Early research gave preponderant attention to oral tradition, which it sought to distinguish from literacy influence.*

This emphasis on oral sources was part of the text-critical strategy for tracing the history of text-type transmission according to which these variants would create an inaccurate impression if treated as conventional of the inherited oral tradition. However, this attitude could have consequences for handling sources, like discarding masses of variants, as was done for example by Jan de Vries (example in Frog 2011: 82–83). The concern is unwarranted when using phylogenetic methods, which analyze taxa as brothers or cousins rather than assessing them as a lineage *per se* (each example is at the top of the stemmatic tree of relations; none are in an intermediate position). Phylogenetic methods infer a lineage based on the proximate relation of many elements at the same level. It does not need to presume a gap between the true folktale and literary adaptations. The effect of horizontal transmission (i.e. if literary adaptations draw on elements from other cultural traditions and only partly reflect inherited culture) has been addressed in an optimistic fashion by Greenhill et al. (2009) and by Curie et al. (2010).

#### 5. *Source-critical problems.*

Criticisms against the HGM in the latter part of the 20<sup>th</sup> century included issues raised by the sources used and source-critical standards. These criticisms were in part associated with changes in source-critical standards more generally (Frog 2013b) but a significant factor in broad comparative research was and remains reliance on edited and translated materials owing to the number of languages accessible to any one researcher. Lévi-Strauss (1958: 232) notes that a mythic message is preserved even through the worst translation. The translation could nevertheless have an impact on the encoding of specific traits for phylogenetic analysis if ‘the worst translation’ alters surface details of images and motifs through which the mythic message is communicated. A selection among the versions used in analysis is therefore necessary. The Classic HGM advocated the principle that analysis should be based on an as extensive and exhaustive a corpus as

possible, but then those materials were assessed and sorted according to contemporary source-critical standards (cf. point 4 above). As noted, the introduction, omission or alteration of elements in e.g. a translation of low source-critical quality may affect results in a phylogenetic analysis. This approach needs to maintain qualitative valuations of individual variants and cannot be purely quantitative, because the validity of the outcome of analysis will be dependent on the quality and representativeness of the data. However, it may be noted that phylogenetic analysis could be used as a tool in a larger corpus to assess the probability that certain traits in variants of low source-critical quality accurately reflect local or cultural tradition, or whether these may have been introduced by a collector/author/redactor.

#### 6. *The decontextualization of sources and presumptions of relatedness*

The decontextualization of sources is normally now thought of in terms of isolation from a performance context. It was problematic in earlier research because sections of text relevant for comparison were frequently cut from their context in more complex narratives. This was particularly problematic in motif analysis but also in tale-type analysis where, for example, certain traits of a tale were clearly outcomes of adapting the narrative to the context of a longer story or integrating it into that plot. Some such comparative analyses *presumed* a historical relationship and thus parts recorded in different tales might even be first combined as a reconstruction of the historical local or cultural tradition for comparison. However, this type of reconstruction presents a hermeneutical problem and such synthetic reconstructions should not be included rather than primary sources in a data-set to be analysed. The issue of decontextualization can then be in part mitigated by the coverage of the maximal amount of text for each example (in the present case, for example, not isolating the motif of the escape from Polyphemus’ cave but also all of the surrounding tale).

#### 7. *The representativeness of sources.*

The problem of the representativeness of sources is a question of whether isolated

examples can be considered representative of a local or cultural tradition. This is particularly relevant to phylogenetic analysis on the basis of individual formal traits. Some simple examples of this are the examples of ATU 1148b attested in Sámi, Latvian and Greek discussed by Frog (2011: 81, 84, 87). This is particularly significant for the types of interpretations discussed when different variants of a tale from a single cultural group do not systematically group together as more closely related to one another than to those of other groups (cf. the distribution of Sámi and Greek/Homeric variants in Fig. 1). This problem requires a close analysis to establish whether the variants present different locally established forms, which could be born/borrowed at different times, or if a local teller know both the traditional and an anomalous tale at the same time. Concern over whether an example is historically rooted in one culture as opposed to borrowed through contacts with another may be alleviated when focus is calibrated to a broader scope: for example, it becomes unnecessary to resolve whether a Sámi example reflects a borrowing from Russian or Norwegian tradition if comparison is between European/Eurasian traditions and traditions in the Americas and individual examples are considered in relation to those broad patterns (cf. below).

#### *8. Researcher interpretation in type-identification.*

The researcher's identification of an example or group of examples could be inaccurate or irrelevant, such as the Sámi examples reviewed in Frog (2011: 81) that are identified with ATU 1148b on the basis of the historical reconstruction of their relatedness to the abstract tale-type rather than purely on the basis of formal features of the individual examples. If this sort of identification is considered justified, it is then followed by the problem that many similar cases remain unidentified and the additional problem that such loose groupings may not in all cases be valid. This becomes a problem of hermeneutics: to what extent does looking for parallels produce parallels and their justification? In the background of this question appears to be the criticisms of Kaarle

Krohn's (1926: 28–29) conception that each motif has a single unique origin, which rejects the possibility of 'multigenesis' of narrative elements (cf. Frog 2013b: 27, 31n.13). This is a very controversial issue that could be statistically evaluated for each motif thanks to statistical tools, for example by estimating how many founder events are necessary to explain the diversity of a studied corpus. A solution may also be to search for a sufficiently complex set of traits that could not be the product of many independent inventions around the world. The researcher's identification of motifs / elements of the text may nevertheless remain a problem. This problem is similar to the issue of producing parallels by looking for them: what qualifies as presence/absence or 'the same'/'different' remains dependent on researcher interpretation, and this is complemented by the problem of researcher subjectivity in determining which elements are relevant for observation and which are not. A solution could be to determine the maximal number of elements for each text subjected to analysis. It should also be noted that varying the number and categorisation of elements subject to analysis often does not change the overall result (d'Huy 2013c; 2013f).

#### ***Phylogenetic Analysis of the Polyphemus Tale***

The reconstruction of the Polyphemus tale is a textbook case. The earlier reconstructions of the proto-myth, and of the significance which lies at the root of the story, can be safely dismissed as erroneous. This is a broad subject, too wide to be reviewed here, and the reader may consult Justin Glenn (1978) for an introduction. The most complete attempt to reconstruct the proto-version of Polyphemus was O. Hackman's analysis based on a Historical-Geographical approach (Hackman 1904). This study suffers from a total lack of explanation for the criteria used to limit the number of versions included in the corpus (Calame 1995: 143). The problem of the physical, geographical origin of this story also seems unsolvable (Glenn 1978).

I have previously applied phylogenetic methods to the historical reconstruction of the Polyphemus Tale elsewhere (d'Huy 2012a;

Table 2. Examples and sources used in the phylogenetic analysis.

Language / Language Family	№ of Variants	Sources
Algonquian	4	Ojibwa people (Desveaux 1988: 83) Atsina people (Kroeber 1907: 65–67) Niitsitapi people (Spence 1914: 208–212; Wissler & Duvall 1908: 50–52)
Iroquoian	3	Crew people (Lowie 1918: 216–217, 218–220; Simms 1903: 295–297)
Southern Athabaskan	5	Jicarilla Apache people (Goddard 1911: 212–214; Opler 1938: 256–260) Kiowa Apache people (McAllister 1949: 52–53) Lipan Apache people (Opler 1940: 122–125) Chiricahua Apache people (Opler 1942: 15–18)
Tanoan	2	Kiowa people (Parsons 1929: 21–24, 25–26)
Greek	4	Homer, <i>The Odyssey</i> (book IX) Modern Greek people (Athens: Drosinis 1884: 170–176; Cappadocia: Dawkins 1916: 551; Chios: Ludwig 1863: 287–289)
Albanian	1	Albanian people (Comparetti 1875: 308–310)
Italic	10	Abbruzzian people (Nino 1883: 305–307) Sicilian people (Crane 1885: 89) Jean de Haute-Seille, <i>Li romans de Dolopathos</i> Gascon people (Bladé 1886; Dardy 1884) Romanian people (Grimm 1857: 15–16) Valais people (Abry 2002: 58)
Balto-Slavic	3	Serb people (Karadschitsch 1854: 222–225, Krauss 1883: 170–173) Russian people (Ralston 1873: 178–181; Karel 1907: 38–39) Lithuanian people (Richter 1889: 87–89)
Germanic	2	English people (Baring-Gould 1890) West Highlands people (Campbell 1860: 105–114).
Indo-Iranian	1	Ossetian people (Dirr 1922: 262)
Caucasian	2	Abaza people (Colarusso 2002: 200–202; Dumézil 1965: 55–59)
Uralic	3	Hungarian people (Stier 1857: 146–150) Sami people (Poestion 1886: 122–126; 152–154),
Kartvelian	1	Mingrelia people (Frazer 1921: 449–450)
Turkic	2	Oghuz Turks people ( <i>Book of Dede Korkut</i> ) Kyrgyz from Pamir (Dor 1983: 34–36)
Afro-Asiatic	6	Berbers (Germain 1935; Frobenius 1996: 38–41) Palestinian-Israelian people (Patai 1998: 31–32) Syrian people (Prym et Socin 1881: 115)
Language isolates	5	Kootenays people (Boas 1918: 213–219, 303–304) Basque people (Cerquand 1992; Vinson 1883: 42–45; Webster 1879: 4–6)

2013a) and, in the preceding issue of this journal, I used a corpus of examples of this tradition to explore the potential of Natural Language Processing software for identifying motifs (d’Huy 2014c). My first preliminary attempts to reconstruct the evolution of Polyphemus faced major problems owing to the initial sample sizes (24 versions analysed according to 72 traits in d’Huy 2012a; 44 versions according to 98 traits in d’Huy 2013a). I here increase the number of versions

(56) and traits (190) studied. In this paper, I will test my earlier results.

Stith Thompson (1961) counted five traditional elements or motifs in Polyphemus tale-type: G100: Giant ogre, Polyphemus; K1011: Eye-remedy. Under the pretence of curing his eyesight, the trickster blinds the dupe (Often with a glowing mass thrust into the eye); K521.1: Escape by dressing in animal (bird, human) skin; K602: “Noman”; K603: Escape under ram’s belly. Uther (2004)



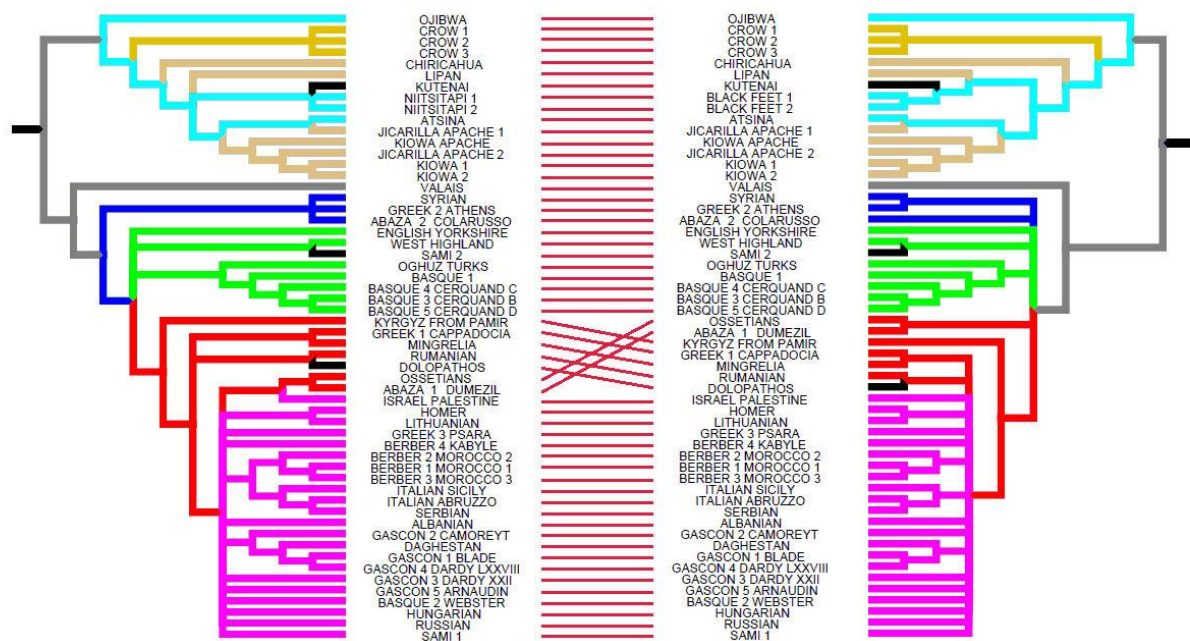


Figure 1. Tree under the maximum parsimony and consensus criterions (left) and bayesian tree (right).

adds five additional motifs: F512: Person unusual as to his eyes; F531: Giant; K1010: Deception through false doctoring; K521: Escape by disguise; D1612.1: Magic objects betray fugitive. Give alarm when fugitive escapes. These motifs can be found in disparate ways in other tales, and each of them has its own evolutionary story. So, in this study, I will only consider the motif of the escape from Polyphemus' cave (K521; K603) and I define the Polyphemus type as a tale in which a person gets into the homestead of a master of animals or of a monstrous shepherd; the host wants to kill the hero, but the hero escapes by holding on to the fleece or fur of an animal who is going out, concealing himself under an animal's skin or with a living animal.<sup>2</sup>

The versions are drawn from diverse published sources in several languages (English, French, German, Italian). Some of the sources used were not available in forms that are up to modern source-critical standards and may have potentially been subject to significant editing for the earlier publication or could reflect summaries and paraphrases (although see discussion above). The present study is founded on the premise that the texts forming the corpus are sufficiently representative of the traditions of the cultures in question to make phylogenetic analysis reasonable. This also means that the reliability

of the results remains conditional on the representativeness of the corpus.

Each version of the Polyphemus Tale has been analysed individually, breaking it into the shortest possible sentences. These sentences have then been added to an index to compare the mythological versions they contain. The sentences were coded according to their presence in (1) or absence (0) from each version, in order to produce a binary matrix. The coding also incorporated a symbol (?) for uncertainty in the data.

With Mesquite 2.75 (Madisson & Madisson 2011), a simple model to calculate the 100 more parsimonious trees was used. Then each tree was rearranged by subtree pruning and regrafting, before being summarized into one – consensual – tree (strict consensus; treelength: 608; Figure 1, left column). With MrBayes 3.2.1 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003), the posterior distribution of phylogenetic tree for all the versions was inferred. An ordinary Markov Chain Monte Carlo analysis for 20,000,000 generations with 4 chains was run, using a model of DNA substitution (the GTR) with gamma-distributed rate variation across sites. The trees were sampled every 5,000 generations, with relative burn-in discarding the first 25% of sampled trees. The fact that a stationary distribution of values had been reached was controlled with Tracer 1.5.0 (Drummond &

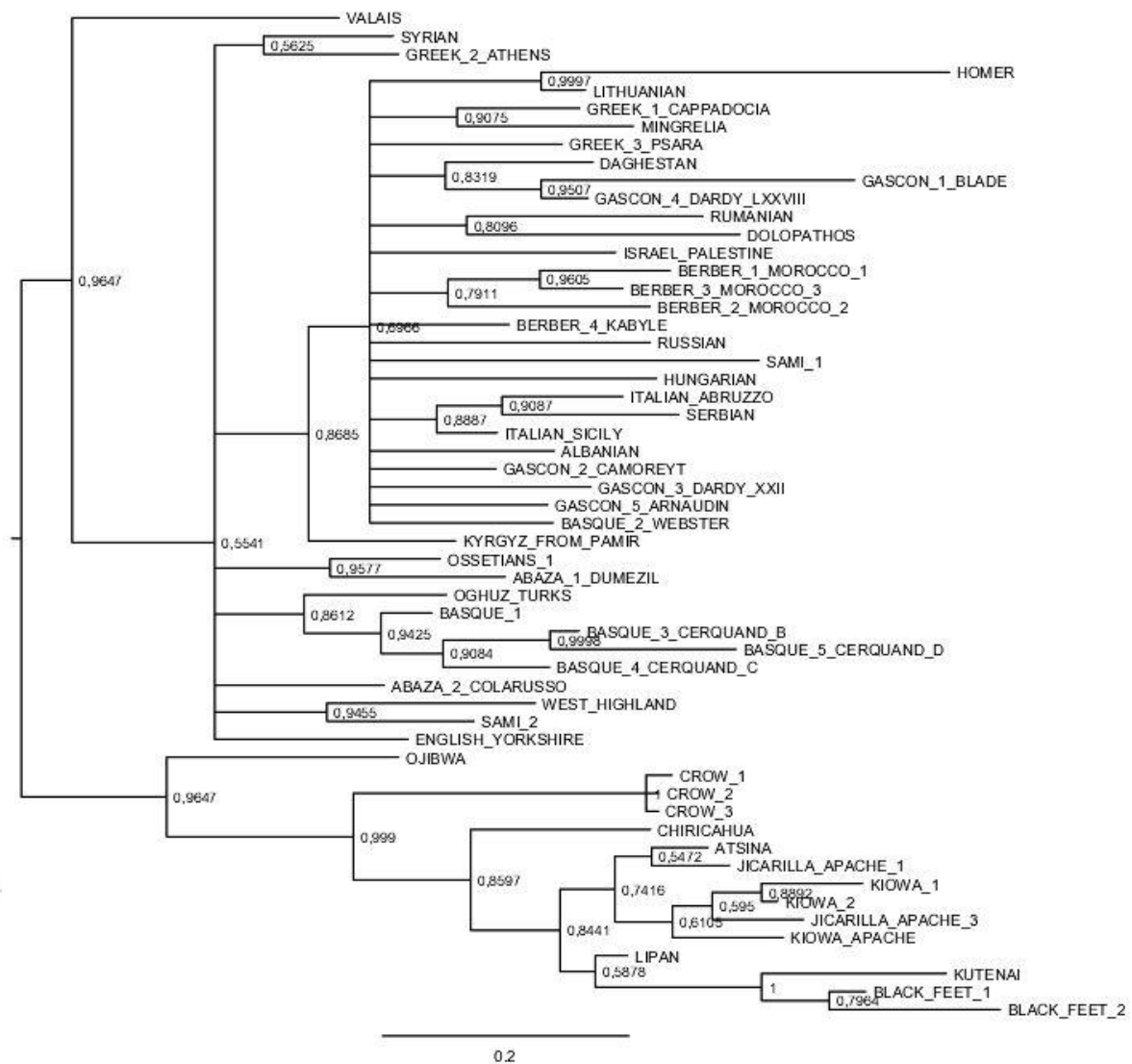


Figure 2. Maximum tree with a midpoint rooting.

Rambaut 2007). At the end of the run, the average standard deviation of split frequencies was 0.005. Both runs produced 8,002 trees, of which 6,002 were sampled. The tree obtained is a consensus tree from all samples (excluding the burn-in), created by a 50% majority rule. This means that a polytomy is introduced if a particular split occurs in less than 50% of all trees and so the program was unable to resolve this lineage (Figure 1, right column).

To root the trees, I used a midpoint solution with the MrBayes tree, which places the root directly between the Ojibwa and Valais versions (Figure 2). The phylogenetic link between both versions possess a very strong confidence degree (0.97) and was systematically found in the previous reconstructions (d'Huy 2012a; 2013a). On the one hand, the Valais is formally intermediate,

between the European and Amerindian corpora, with a lord of wild animals similar to Amerindian versions found in the Valais corpus. It is likely that the European version exhibits the most archaic features.

Considering the monster in the earliest shared form of the tale as a lord of animals, as in the Valais<sup>3</sup> and North American variants, would be in agreement with Burkert's statement (1979: 33) that the Cyclops in Homer drew on a primeval mythological tradition older than the Indo-European tradition that included a belief in a lord of animals. As pointed out by Frog (p.c.), narrative traditions and the images of different categories of imaginal being adapt and are shaped historically in relation to dominant livelihoods of the cultural environment in both legends and mythology (cf. af Klintberg 2010: 168). The supernatural shepherd of

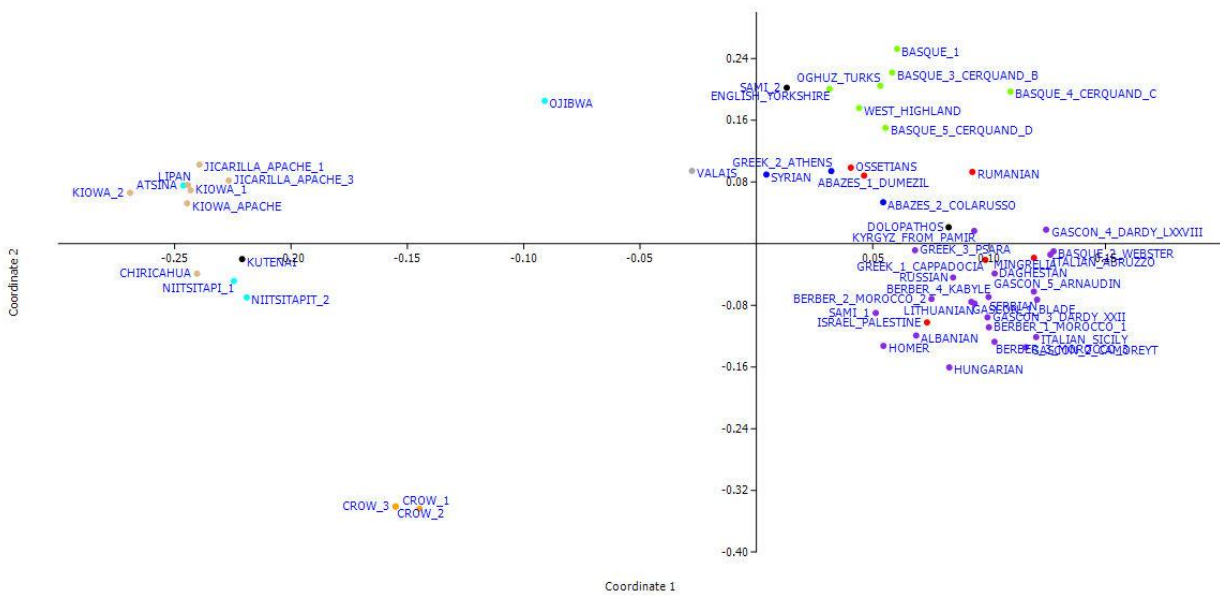


Figure 3. Principal coordinates analysis (Jaccard).

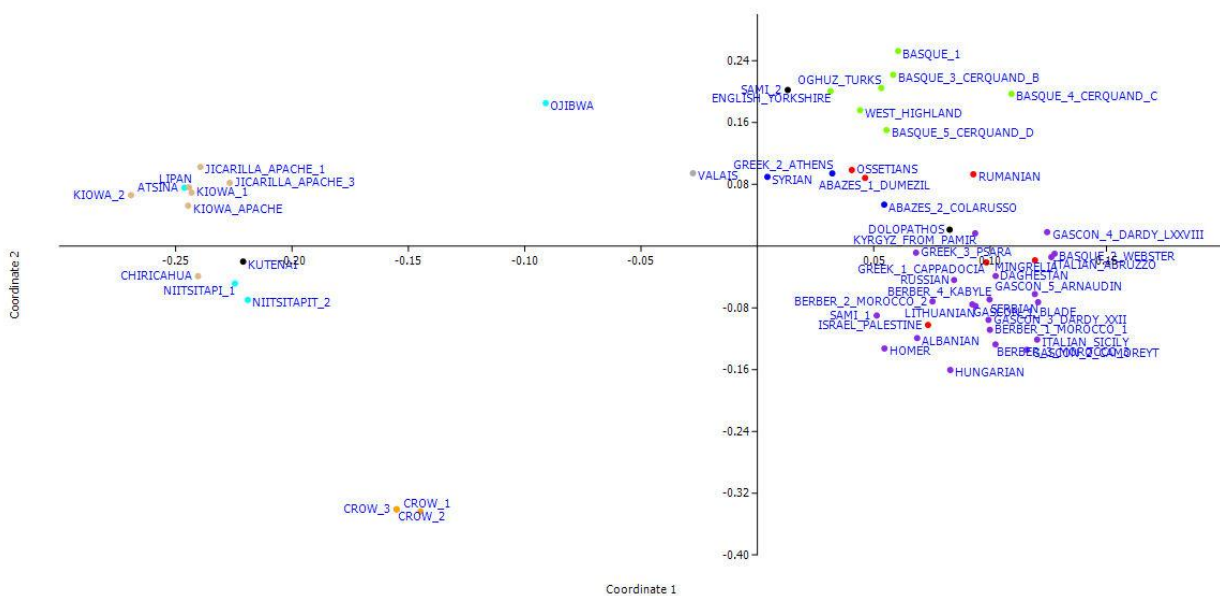


Figure 4. Principal coordinates analysis (Cosine).

European traditions is equivalent to the lord of animals in his control of resources while the resources concerned are connected to different kinds of livelihoods (cf. also Tolley 2012, which discusses a motif associated with the lord of animals also adapted to livestock).

The plot of the Polyphemus tale is structurally dependent on the monster being a keeper of animals, on which the hero's escape is dependent, and which would account for its long-term stability as an element of the plot

(see also Frog 2011: 91–93; 2014). It is therefore probable that this feature of the tale was established already in the form from which the attested versions derive. *If* the North American and European versions of the narrative are historically related *and* the narrative was not carried to the Americas by late medieval colonization by Europeans,<sup>4</sup> *then* it is improbable that the necessary contact and exchange relevant to the spread of the European version with sheep antedated

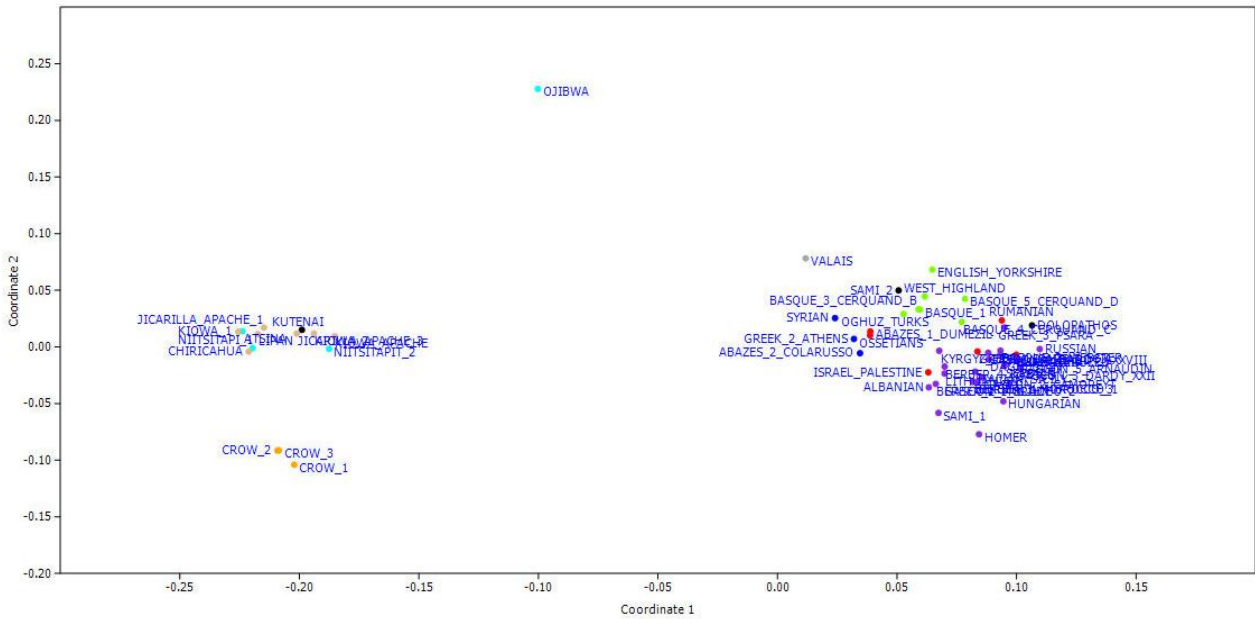


Figure 5. Non-Metric Multidimensional Scaling (Jaccard).

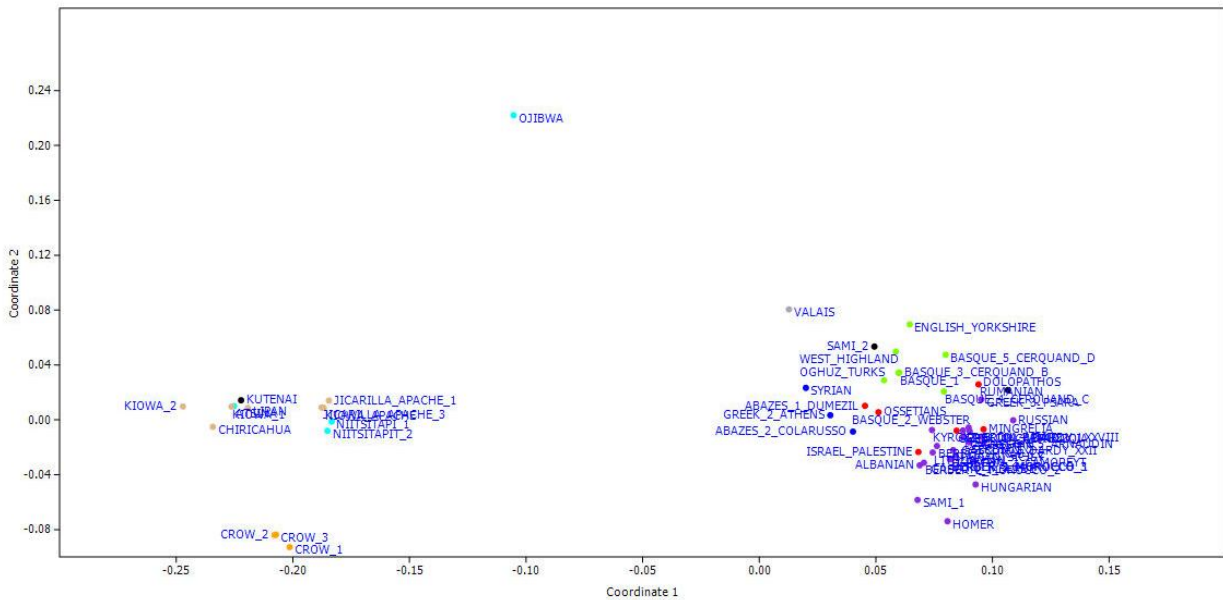


Figure 6. Non-Metric multidimensional Scaling (Cosine)

the domestication of livestock. Accordingly, the adversary was most likely a lord of animals or equivalent figure in the earliest construable form of the plot.

The lord of animals is attested among several peoples in Europe. It is therefore unclear why it would be maintained only in the tradition area of Valais where it is attested in only one variant. The appearance of a lord of animals in the Valais instance may not reflect a historical continuity from such an era before the domestication of livestock that was maintained in isolation in Western Europe. Yet the local evolution of this tale shaped it like the (Palaeolithic) proto-form, which

explains its place in our analysis. On the other hand, the Ojibwa's branch is also situated in an intermediate place. A principal coordinates analysis (transformation exponent:  $c = 2$ ; Similarity index: Jaccard; PC1: 29,859; PC2: 10,07; Cosine: PC1: 35,62; PC2: 11,74; fig.3 and 4) and a non-metric MDS (Jaccard, Cosine, 2D; Figures 5 and 6) conducted with Past 3.0 (Hammer et al. 2001) show a remarkably consistent pattern, geographically speaking (North America / Europe; nearest geographical versions tended to form sister clades), and confirm the intermediary situation of the Valais' and Ojibwa's versions

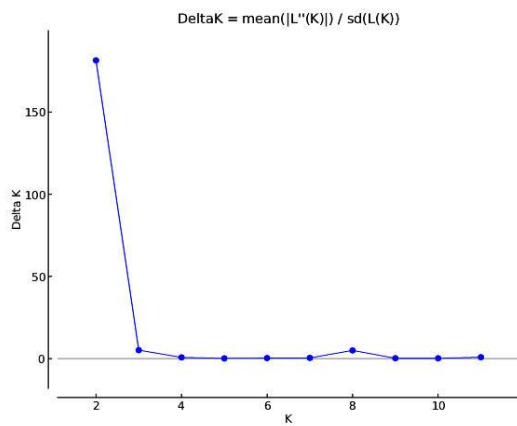


Figure 7. Delta K's score associated with 1 to 12 clusters.

as exhibiting formal distinction from these larger groups. I also used Structure2.3.4 (Pritchard et al. 2000; Falush et al. 2003) to detect the true number of clusters (K, test for 1 to 12 clusters) in the sample of versions studied. Using the software structure Harvester (Earl & von Holdt 2012), two main clusters are identified among the variants in the way that the variants within a cluster are more similar to each other than to the other cluster (Parameters: 10,000 Burn-in period; 50,000 MCMC Reps after burn-in; number of iteration: 10; recessive alleles model used for

0; Ancestry Model; Admixture model; Figure 7). The data align perfectly with the Amerindian/European distinction.

The software also computes the probabilities of each version for each cluster. The probability is by far the lowest for the Ojibwa (0.53% for the Amerindian cluster; 0.47% for the European cluster) and the Valais (0.28% for the Amerindian cluster; 0.72% for the European cluster); this again suggests that these two versions are in the middle ground between European and Amerindian developments. The limited number of examples from each culture in the corpus may not be sufficient to reconstruct the conventional form of the tradition for any one culture in a dependable manner. Nevertheless, the phylogenetic analysis clearly shows distinct groupings of the European and North American branches of the tale. Although the historical background behind the branching of the Ojibwa (as well as the Crow) and Valais examples is unclear, it remains noteworthy that significant formal variations in the European and North American clusters are inclined toward the center of shared features of the traditions rather than away from it at random. This makes it appear less likely that



Figure 8a. The 'Petit Sorcier à l'Arc Musical' ['The Sorcerer with the Musical Bow'] in the Cave of the Trois-Frères in Ariège, southwestern France, Magdalenian, may be the earliest pictographic representation of the Polyphemus tale (Breuil 1930: 262).

the two major branches of this complex narrative emerged independently of one another.

#### ***An Example from Palaeolithic Rock Art?***

An illustration of the Polyphemus tale can potentially be interpreted from the Palaeolithic cave drawings found in the Trois-Frères. This cave is located in Montesquieu-Avantès, in the French Ariège *département* and the cave drawings appear to date to the Magdalenian period, long before the first domestication of animals.

The potential case is included as a scene within a dense, superimposed and complex representation of a herd (Figure 8a). The scene in question depicts a bison-man with a bow in his hand (on which see further Demouche et al. 1996). This figure is striking in that it appears to be a rather detailed representation of a bison standing on its hind, human, legs and holding or pointing a bow. This being observes one of the animals which – if correctly interpreted – has a human thigh (see discussions in Breuil 1930: 263; Leroi-Gourhan 1971: 97) and a very detailed, large anus / vulva (Breuil 1930: 261; Vialou 1987: 116), as seen more clearly in Figure 8b.

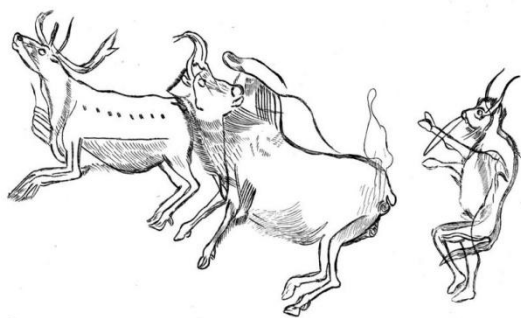


Figure 8b. The images of Figure 8a that may be relevant to the Polyphemus tale (Breuil 1930: 262; Breuil's drawing).

Interpreting such images is necessarily speculative and problematic. A popular interpretation is that the figure of the human-bison is a 'shaman'. The bison-man could be interpreted as some type of magical hunter, but the bison-man head identifies him with the herd of animals and suggests his identity is somehow connected to the herd by the features he has in common with it, rather than those that are different from it. Some believe that they represent hunters in animal disguise (Demouche et al. 1996), in a way similar to

the one used by the North American Lakota hunters approaching their prey. The 'Petit Sorcier à l'Arc Musical' has also been described as a man with a bison head playing an instrument, a flute or a musical bow (Bégouën & Breuil 1958: 58). Another possibility is that this figure is not separated from the herd as a hunter or predator but rather aligned with them as their protector, guardian or other agent and representative (Clottes & Lewis-Williams 1996: 94).

The peculiar image of the animal with a human thigh and prominent rear orifice is equally obscure, but can be compared to the Amerindian versions of the Polyphemus tale, in which the hero often hides inside an animal itself by entering through its anus. This enables the hero to escape the monster who controls the beasts from his dwelling. A motif of the hero hiding in this way would account for the prominence of the anus / vulva on the depicted animal and the co-occurrence of this with the peculiar feature of a human thigh on the animal. In addition, it would also account for the relationship to the upright bison-man looking at the animal within the context of a herd: the bison-man would then fill the role of a supernatural guardian of a herd watching for the hero who escapes by hiding within one of the animals.

Interpreting narrative through image systems of a remote earlier period is inevitably problematic and speculative. If this set of image elements indeed belongs together, they can reasonably be presumed to reflect some sort of a narrative through its constituent elements. The narrative depicted might be random, local or reflect an imaginal depiction of a historical event, but its choice as a subject for representation could also be connected to some type of social prominence or relevance. Comparative evidence supports the probability that the Polyphemus tale was current in some form in the Palaeolithic era, and its *longue durée* is a relevant indicator that it held social interest and relevance. Provided that the set of image-elements have been more or less accurately interpreted, they would appear to parallel elements that stand at the core of the Prometheus tale – i.e. the escape of the hero. The bison-man would also be consistent with the proposed evolution of

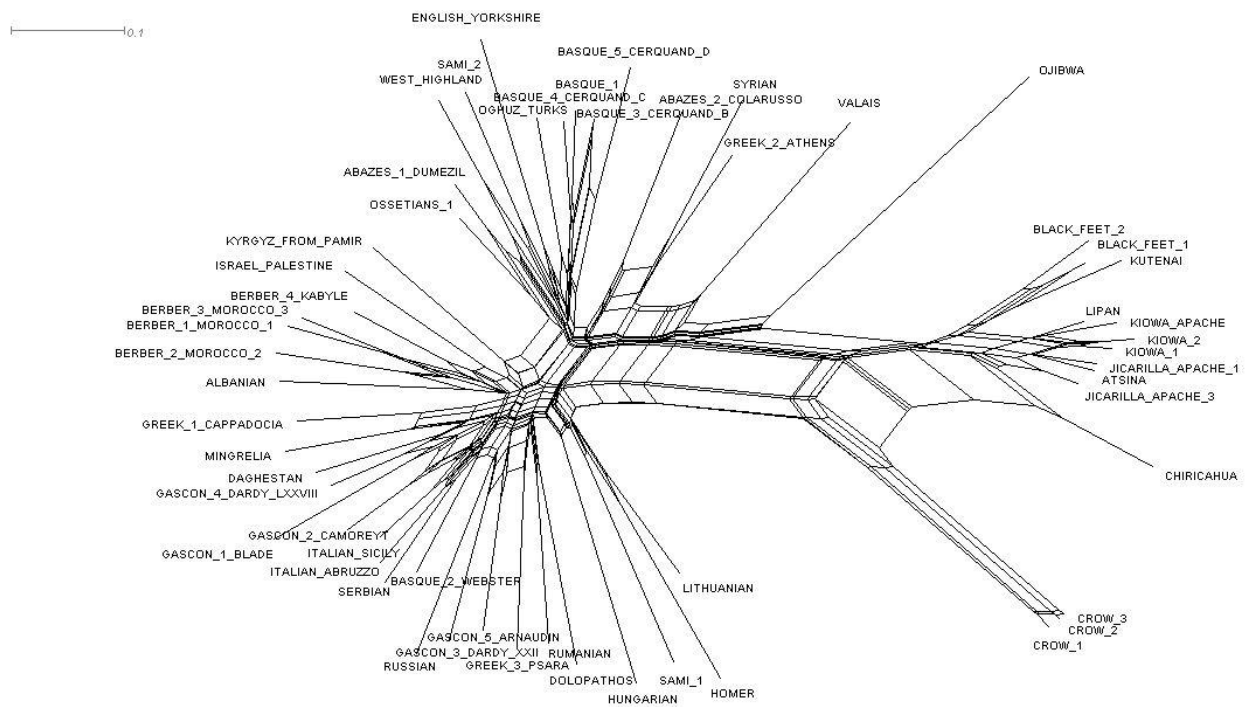


Figure 9. NeighborNet graph of the Polyphemus variants.

the tale's protagonist in Europe/Eurasia from a guardian of animals into a herdsman of domesticated livestock (noting that here he may be guardian of a particular species, notably a herd animal). This interpretation is speculative, but it is not unreasonable and is worth putting forward owing to what we know of the tale and can infer about its history.

### **The Tale's Retention Index**

If Polyphemus is a Palaeolithic tale, then, in the model of its history, we would expect the rate of borrowing of mythemes to be low. To test this, the Retention Index (RI) has been calculated for both our trees. The Retention Index is a traditional tool in cladistics and evaluates the degree to which a trait is shared through common descent. An RI of 1 indicates that the tree shows no borrowings, while an RI of 0 indicates the maximum amount of borrowings that is possible. The RI calculated with Mesquite was 0.57 for the Mesquite tree and 0.63 for the Bayesian tree (Jukes-Cantor model; 1000 characters simulated). These indices indicated that most of the mythemes were shared through common descent. Indeed, high RI values (for instance, greater than 0.60) usually show a low horizontal transmission (Nunn et al. 2010). Both RIs obtained (0.57 and 0.63) are

broadly the same as the mean RIs for the biological data sets presented by Collard et al. (2006), whose mean RI is 0.61. The biological data sets of Collard et al. were structured by speciation. Thus the vertical transmission (from mother to daughter populations) should be the dominant evolutionary process in both biological and folktale data. However, note that the RI for the Polyphemus myth does not look sufficiently high enough to consider it completely significant rather than, for example, explaining it as an interpretive bias in selecting, labelling or interpreting data.

The results also should be controlled with NeighborNet (implemented in Splitstree4.12; Bryant & Moulton 2004; Bryant et al. 2005; characters transformation: Jaccard; Figure 9). This algorithm makes it possible to see conflicting data, noise, doubt, uncertainty as webbing, and proposes good representations about both clusters and evolutionary distances between the taxa. A real conflicting signal between versions (box-like structures) was found. However, NeighborNet correctly brings the major part of these versions together into coherent geographical or cultural clusters, similar to those found in both trees (see below), suggesting a good conservation of the stories. The main delta-score is here 0.3422. The delta-method scores individual

taxa from 0 to 1; a relatively high delta score (close to 1) shows a strong conflicting signal in the data (Holland et al. 2002). Whichmann et al. (2011) calculates delta scores across the world's language families. Their average is 0.3113. Thus, contrary to Ross and al.'s claims (2013), some families of folktales, including K603, are at least as tree-like as languages, if not more so.

### ***The Reconstruction of a Protoversion***

Two phylogenetic comparative methods (Maximum Likelihood with model Mk1 and Parsimony reconstructions) implemented in Mesquite 2.75 have been used with maximum parsimony and consensus criteria to reconstruct the probable form of the first Palaeolithic state of the Polyphemus family. These phylogenetic reconstruction methods are applied to each mytheme of the family. Then mythemes reconstructed with a high degree of confidence – i.e. with more than a 50% probability using both methods – have been retained. In the text, mythemes with more than 75% probability have been underlined.

[The enemy is a completely solitary figure, who is affronted alone.] A human hunter enters in the monster's house [which is a hut, a house or something similar]. [The hero does not know whom he will meet.] The monster possesses herd of wild animals. [He traps the man and his own animals with an immovable or a large door.] Then he waits for the man near the entrance to kill him. [To escape, the hero clings to a living animal.] In this story, a vengeance occurs that is connected with fire.

This abstract is very close to what has been found previously (d'Huy 2013a) using fewer versions and another choice of traits to study the tale. It could be the Palaeolithic myth of the first appearance of game on Earth. Phylogenetic methods cannot discover the original form of a story in the sense of an Urform with certainty, yet they can propose statistical reconstructions, where reconstructed traits are not necessarily those which occur most frequently. Note that this model is linked to features that are also correlated with the Amerindian traditions.

We can attempt to correlate the trees with a model of historical spread. However, this is a very hypothetical reconstruction. In Europe, the palaeolithic populations may have migrated toward the South (fig. 3, in blue) during the Last Glacial Maximum (Pala et al. 2012; Peričić et al. 2005) and probably preserved at least partially a reconstructed version of the story in which the monster was a master of animals. If the Valais variant is left aside as an exceptional outlier in the data and the branching of the Syrian, Greek 2 and Abaze variants' cluster treated as the root point of the European variants' stemma (fig.3, tree at the left), the following text, which may approximate the features of versions of the European Neolithic proto-tale, can be reconstructed:

The enemy is a completely solitary figure, a giant who has one eye in the forehead, and is affronted alone by the hero. A human [perceives a light in the distance and does not know whom he will meet]. He enters the monster's house. The monster possesses a herd of domestic animals (sheep). [He traps the man and his own animals with an immovable or a large door.] Then he falls asleep and a vengeance occurs that is connected with fire. The monster waits for the man near the entrance to kill him. [To escape, the hero clings to a living animal.]

According to the reconstructed origin of the European type (Figure 3, Greece / Syria / Abaze, in blue), this new version where the monster was in a shelter and the animals were sheep may go back to about the domestication of the species. Indeed, the domestication date of sheep is estimated to fall between nine and eleven thousand years ago in Mesopotamia. If the new Polyphemus' tale type was linked to the early stages of animal domestication, it may have been disseminated through successive migrations from the Mediterranean area across millennia.

This model has been tested by removing the Amerindian data: the Bayesian tree remains almost the same (Figure 10). This makes the outcome appear relatively consistent with what would be developed from stemmatic models developed by other means, because the whole branch is stable (cf. a stemma for Germanic languages should appear more or less the same even if we were



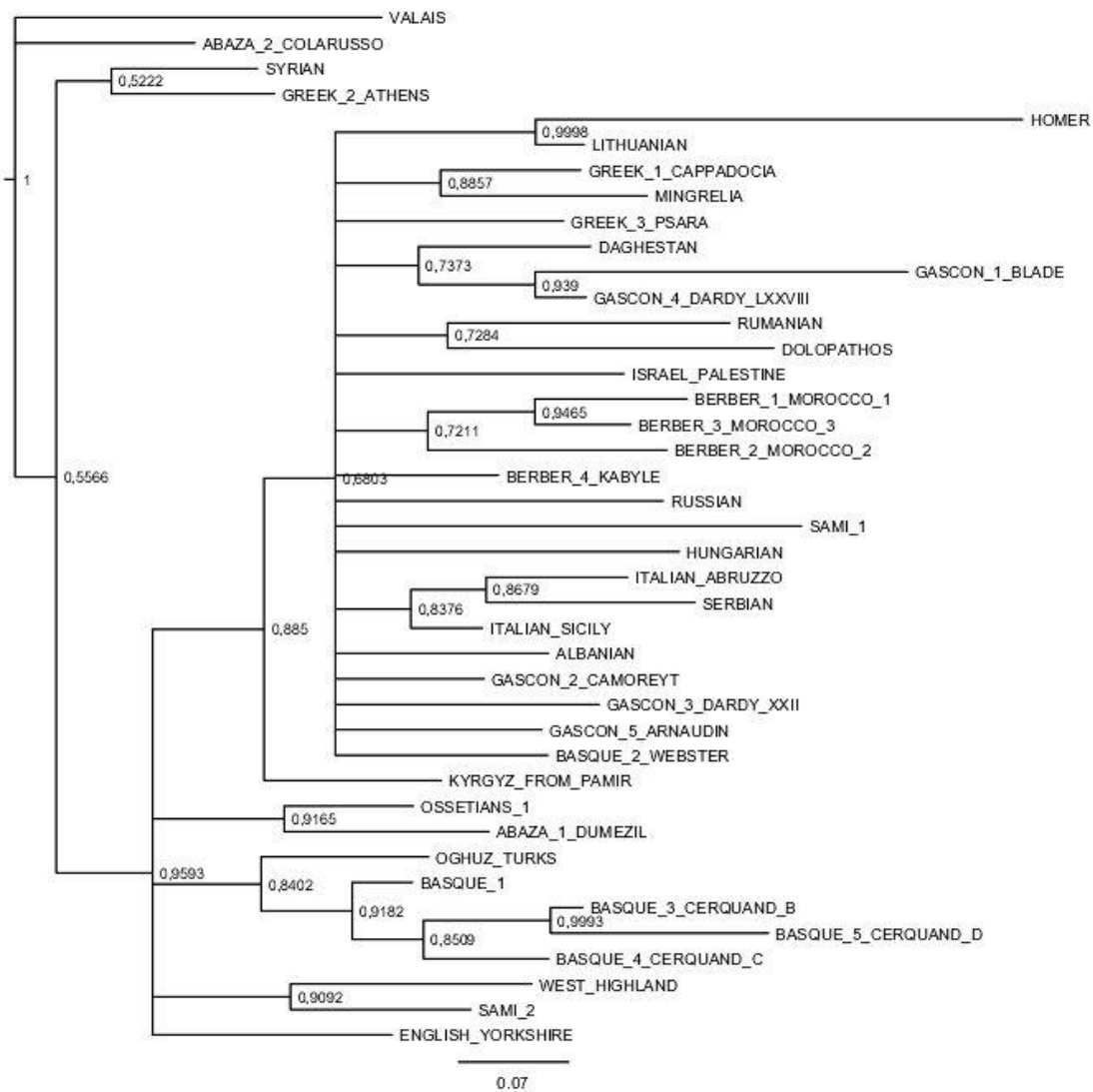


Figure 10. Bayesian tree calculated without the Amerindian versions.

unaware of a connection to Indo-European). Yet this tree alone would not resolve which features in the primary split should be considered probable for an antecedent form other than those shared across that split. For example, Burkert's hypothesis that an earlier form of the tale incorporated a belief in a lord of animals requires the Amerindian branch of data in order to advance beyond speculation to have empirically based support, conditional on the improbability of multigenesis.

### ***Trends of Stability and Contrasts***

Following the working hypothesis that complex narratives of the escape from Polyphemos are unlikely to emerge independently of one another, these stories could only have spread across Eurasia and North America when a former land bridge joined present day Alaska and eastern Siberia during the Pleistocene ice ages. In this case, it becomes necessary to account for the tale-

type's distribution in two very large areas that are geographically remote from one another and diversity within these different areas which only partly seems to correlate with cultural and population histories.

A 10,000 year model of population movements and cultural changes have probably had transformative effects on traditions across Europe and America for millennia. For instance, one can propose that one of the first steps of diffusion in Europe includes Basque, Oghuz Turks, Yorkshire and the West Highlands. I have observed a similar cluster previously (d'Huy 2013) with the use of other mythemes to study the Polyphemos' tales. It could be easy to explain. During the first millennium BC, Celtic languages were spoken across much of Europe, including Great Britain, the Pyrenean area, the Black Sea and the Northern Balkan Peninsula. The Basque versions may be a borrowing from the neighbouring Celtiberian (spoken in ancient

times in the Iberian Peninsula) or Gaulish languages. Yorkshire belongs to the Brittonic area, and the West Highlands is included in Goidelic. The link between the Pyrenean area and Oghuz Turks could be explained by the Gallic invasion of the Balkans in 279 BC. More precisely, the Tectosages, one of the three tribes who settled Galatia (an area in the highlands of central Anatolia) ca. 270 BC, came from southern France and could potentially be the vector of transmission. However, an account of the Celtic establishment of a branch of the tradition could not be shown to be 'true', but as a possible but indemonstrable explanation that would be the outcome of the effect of population movements and cultural changes of traditions. If this has happened repeatedly, it would suggest that different versions of the story have been 'seeded' through Europe again and again, superseding one another and receding in the wake of history. This would consequently seem to make it difficult to correlate the earliest, palaeolithic reconstructable version of the tale with any particular geographical space.

To test the multiple migration hypothesis, I realised a Mantel test using a Jaccard's coefficient matrix (permutation: 10000) on individual version data with SAM v.4.0 (Rangel et al. 2010). If correct, there should be low relationship between geographic distance and similarity between versions, each new version taking the place of older versions, breaking the continuity of linear diffusion. The geographical locations of each version were estimated using information included in the books and papers. I adopted the centroid of geographical coordinates for each language area when no precise geographical information was available (using the websites Glottolog and Wals). I found that geographical distance explains 7% of the variance ( $r^2 = 0.07$ ;  $p = 0.043$ ) in the Amerindian data and 0.8% ( $r^2 = 0.008$ ;  $p = 0.3$ ) among European data. The correlation coefficient detects only linear dependencies between two variables, so this low result suggests a very complex evolution for the European versions, with many waves of diffusions (rather than a single one), and the long-distance influence of certain versions,

such the Homeric one, could explain the clade joining closely Israeli, Berber and Russian versions (see Ross et al. 2013 for higher results about a European folktale). In these conditions, the result may imply that the diffusion of versions could be more phylogenetic (only the existence of a parental version needs to be taken into account) than geographical. Another hypothesis could be a very good conservation of the structure of the tale, which would be borrowing without major modifications.

How could the Polyphemus' tales – and other tales – evolve and survive from the Palaeolithic period? Biology may propose a model (d'Huy 2013a; 2013c–d). The theory of punctuated equilibrium states that when significant evolutionary change occurs in a species, it is generally restricted to rare and very fast events of branching speciation (Eldredge & Gould 1972; Gould & Eldredge 1977). If an analogy may be drawn, newly mythological sister versions would tend to diverge rapidly, which would be followed by extended periods of stability with little net evolutionary change, or what Frog (2011: 91) has described as “the evolution of tradition [...] in fits and starts.” One sign of the punctuational evolution of myths is the correlation between branch length and the number of speciation events (Webster et al. 2003; Pagel et al. 2006). Where many speciation events (nodes) have occurred, there should be more total genetic change (longer path lengths). A gradual model of evolution predicts no relationship between node and path lengths.

The mean length has been calculated for each branch of the Bayesian tree (Figure 2) from the final version to the first polytomy – more than two based branches, which is also a sign of punctuational evolution (Wagner & Erwin 1995), and not necessary from the root of the tree to each final version. The mean linear relationship (Pearson + Spearman<sup>5</sup>) square between path lengths and nodes in the MrBayes' tree has been used to give an estimate of the punctuational effect on the clock-like behaviour of these trees. The result was 0.85 (Pearson: 0.91;  $p(\text{uncorr})$ : 2.15E-22; Spearman's  $r_s$ : 0.79;  $p(\text{uncorr})$ : 6.77E-13). The results have been far superior to those

obtained from biological data ( $r$  = from 0.22 to 0.69; mean  $R^2$  = 0.18; Pagel et al. 2006), showing a greater change of the tree length attributable to punctuational effects. The remaining variation in path could be explained by independent gradual effects.

A well-known artifact of phylogenetic reconstructions (the so-called ‘node-density artifact’) may lead us to believe in a false punctuated equilibrium effect. To avoid this, the coefficient of determination ( $R^2$ ) has been calculated. An  $R^2$  near 1.0 indicates that a regression line fits the data well, while an  $R^2$  closer to 0 indicates a regression line does not fit the data very well. Here, the  $R^2$  with a linear regression ( $R^2$  = 0.83) is higher than the  $R^2$  with a logarithmic regression ( $R^2$  = 0.75). Trees also did not show the curvilinear trend that characterizes the node-density artefact (Venditti et al. 2006). The punctuational effect for this folktale is stronger than the punctuational effect in biological species (22%; Pagel et al. 2006) or in languages (10–33% being the overall vocabulary differences among languages within a language family; see Atkinson et al. 2008). It should contribute 75% to the evolution of the Polyphemus tale, a result close to what was obtained for another tale-type: the Cosmic hunt (84%: d’Huy 2013c: 100).

Ethnology provides a model that could explain these mythological punctuations. Folktale variations are largely defined by people drawing a line between ‘us’ and ‘them’ (Ross et al. 2013). Punctuation may thus reflect a human capacity to enhance both the group identity and the identification of individuals with this group. For instance, a story of the origin of fire was told by an Amerindian to offset another story by an Indian of another tribe (Goddard 1904: 197), and people belonging to a certain tribe explained that another tribe with whom it shares many myths did not know how to tell them (Désveaux 2001: 85). The punctuational effect also could be due to a mythological founder effect; small social communities tend to lose part of their mythological complex and experience something similar to founder events and drift, which increase the rate of change (for an example, see d’Huy 2013f).

## Conclusion

To conclude, phylogenetic and statistical tools used to study folktale allow us to return to considerations of the past behind the documented evidence. They can offer insights into how a tale evolves, into the tale’s possible prototype, and to what extent the versions studied belong to a same tale-type, with a common ancestor. Concerning the family of this folktale, the trees obtained are better and more coherent than those obtained in previous studies, which shows the importance of experimental replications and using a larger database. The proto-myth reconstruction and the punctuational evolution of the folktale, also found in previous works, have been corroborated here.

*Julien d’Huy (dhuy.julien[at]yahoo.fr) Institute of the African World (IMAF, UMR 8171), Aix-Marseille University, Paris I Sorbonne; (CNRS/IRD/EHESS/Univ.Paris1/EPHE/Aix-Marseille Univ-AMU), Centre Malher, 9, rue Malher, 75004 Paris, France.*

*Acknowledgements: The author gratefully acknowledges comments from Frog that played a significant role in developing this paper.*

## Notes

1. Pygmalion: d’Huy 2012c; 2013a; 2013f; the Cosmic Hunt: d’Huy 2012b; 2013c; Polyphemus: d’Huy 2012a; 2013a; the Dragon: d’Huy 2013e; 2014a; Little Red Riding Hood: Tehrani 2013; the Kind and the Unkind Girl: Ross et al. 2013; d’Huy 2014b; d’Huy & Dupanloup 2015.
2. Traits were selected for the whole tale in order to avoid the possibility that variants could be grouped together in analysis owing to a concentration of formal similarity in the co-occurrence of motif elements in one episode even if the overall narrative form and structure was close to that of another set of variants. Another approach, not used here, could be to use the tools belonging to the field of Natural Language Processing. With these tools, the closer the contents of two narratives (as reflected through their surface texts), the shorter the distance between the narratives would be. This coding would concern the whole text and avoid the pre-selection of traits (which is perhaps not so significant: see d’Huy 2013f). However, such an approach would require taking many precautionary measures (d’Huy 2014c), such as asking which elements should be compared (individual sentences, groups of sentences, parts of text or structural formations) and whether certain words, sentences, paragraphs or the whole text should be rewritten to facilitate the analysis based on the textual surface of a heterogeneously written corpus. It is also necessary to consider how to prevent ambiguity in the identification of unique terms and terms with many

possible significations as well as how original transcriptions and translations, long and short versions, tales collected on site or tales collected from Westernized people under different conditions such as special ceremonies or evening around the campfire (which can influence the content of the tale), be compared if the proximity of elements within a text is a factor. The potential for these tools, partially explored in d'Huy 2013e and d'Huy 2014c, needs to be explored further, but that is a matter for another paper.

3. An initial potential indication that the Valais variant might maintain archaic features appears at the end of the story: the dwarf (structural inversion of the giant) tries to punish the hero by creating an avalanche. Note that, according to old Tyrolian traditions, certain giants protected the singing birds and sheep; they opened the stables for sheep that had been kept indoors too long, set free badly treated cattle and punished cruel people with... avalanches (Rohrich 1976: 142–195). Yet this motif is not strong support for the lord of animals in the Valais story as representing a historical continuity in the form of the protagonist from an *Urform* of the tale.
4. The hypothesis that the tale was carried to the Americas in the late medieval colonization by Europeans is unlikely because of *a*) the coherent clustering of European variants on the one hand and American on the other; and *b*) because of the widespread dissemination of the motif in North America).
5. The Pearson correlation coefficient is sensitive only to a linear relationship between two variables; the Spearman correlation is sensitive when two variables being compared are monotonically related, even if their relationship is not linear. If the variables are independent, Pearson's and Spearman's correlation coefficient is 0; the coefficient is 1 if the variables are perfectly correlated.

## Works Cited

### Sources

- The Book of Dede Korkut*. Trans. Geoffrey Lewis. Harmondsworth: Penguin, 1974.
- Homer, *The Odyssey*. Ed. & trans. Samuel Butler (London: A.C. Fifield, 1900), rev. Timothy Power and Gregory Nagy. Perseus Digital Library. Ed. in Chief Gregory R. Crane. Available at: <http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0218:book=1:card=1>.
- Jean de Hauteseille. *Li romans de Dolopathos*. Ed. Charles Brunet & Anatole de Montaiglon. Paris: P. Jannet, 1856.

### Literature

- Abry, C. 2002. "Omer m'a tuer... ou moi-même en personne". In *La Mythologie et l'Odysée: hommage à Gabriel Germain: Actes du colloque international de Grenoble 20–22 mai 1999*.

- Ed. A. Hurst & F. Letoublon. Genève: Droz. Pp.57–65.
- Atkinson, Quentin D. et al. 2008. "Languages Evolve in Punctual Bursts". *Science* 319: 588.
- Baring-Gould, S. 1890. "The Giant of New Mills, Sessay". *Folk-Lore* 1: 130–131.
- Bégouën, H., & H. Breuil. 1958. *Les Cavernes du Volp: TroisFrères – Tuc d'Audoubert*. Arts et Métiers: Graphiques. Paris.
- Bladé, J.-F. 1886. *Contes de Gascogne I: Contes épiques*. Paris: Maisonneuve & Ch. Leclerc.
- Boas, F. 1918. *Kutenai Tales*. Washington: Smithsonian Institution, Bureau of American Ethnology.
- Breuil, Henri. 1930. "Un dessin de la grotte des Trois frères (Montesquieu-Avantès) Ariège". In *Comptes-rendus des séances de l'année 1930: Académie des inscriptions et belles-lettres* 74(3): 261–264.
- Bryant, D., & V. Moulton. 2004. "Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks". *Molecular Biology and Evolution* 21 (2): 255–265.
- Bryant, D., F. Filimon & R.D. Gray. 2005. "Untangling Our Past: Languages, Trees, Splits and Networks". In *The Evolution of Cultural Diversity: Phylogenetic Approaches*. Ed. R. Mace et al. London: UCL Press. Pp. 67–84.
- Burkert, Walter. 1979. *Structure and History in Greek Myth and Ritual*. Berkeley / Los Angeles / London: University of California Press.
- Calame, C. 1995. *The Craft of Poetic Speech in Ancient Greece*. Trans. J. Orion. Ithaca: Cornell University Press.
- Campbell, John-Francis. 1860. *Popular Tales of the West Highlands Orally Collected I*. Edinburgh: Edmonston & Douglas.
- Cerquand, Jean-Francois. 1992. *Légendes et récits populaires du Pays Basque*. Bordeaux: Auberon.
- Clottes, Jean, & J. David Lewis-Williams. 1996. *Les chamanes de la préhistoire: Transe et magie dans les grottes ornées*. Paris: Seuil.
- Colarusso, John. 2002. *Nart Sagas from the Caucasus; from the Circassians, Abazas, Abkhaz and Ubykhs*. Princeton / Oxford: Princeton University Press.
- Collard, M., S.J. Shennan & J.J. Tehrani. 2006. "Branching, Blending, and the Evolution of Cultural Similarities and Differences among Human Populations". *Evolution and Human Behavior* 27: 169–184.
- Comparetti, Domenico. 1875. *Novelline popolari Italiane*. Rome / Turin / Florence: Ermanno Loescher.
- Crane, Thomas Frederick. 1885. *Italian Popular Tales*. London: Macmillan.
- Currie, Thomas E., Simon Greenhill & Ruth Mace. 2010. "Is Horizontal Transmission Really a Problem for Phylogenetic Comparative Methods? – A Simulation Study Using Continuous Cultural Traits". *Philosophical Transactions of the Royal Society B* 365: 3903–3912.
- Dardy, L. 1891. *Anthologie populaire de l'Albret, sud-ouest de l'Agenais ou Gascogne landaise II: Contes populaires*. Agen: Michel & Medan.

- Dawkins, Richard McGillivray. 1916. *Modern Greek in Asia Minor: A Study of the Dialects of Silli, Cappadocia and Pharasa, with Grammar, Texts, Translations and Glossary*. Cambridge: University Press.
- Demouche, F., L. Slimak & D. Deflandre. 1996. "Nuovelle approche de la gravure du 'petit sorcier' à l'arc musical de la grotte des Trois Frères (Ariège)". *Préhistoire Anthropologie Méditerranéennes* 5: 35–37.
- Désveaux, E. 1988. *Sous le signe de l'ours: Mythes et temporalité chez les Ojibwa septentrionaux*, Paris: La Maison des Sciences de l'Homme,
- Désveaux, E. 2001. *Quadratura Americana: Essai d'anthropologie lévi-straussienne*. Chên-Bourg: Georg Editeur.
- Dirr, Adolphe. 1922. *Kaukasische Marchen*. Iena: Verlegt bei E. Diederichs.
- Dor, Rémy (ed.). 1983. *Contes Kirghiz de la steppe et de la montagne*. Paris: Publications Orientalistes de France.
- Drosinis, Georgios. 1884. *Ländliche Briefe von Georgios Drosinis: Land und Leute in Nord-Euböa*. Trans. A. Boltz. Leipzig: Friedrich.
- Drummond, A.J., & A. Rambaut. 2007. "BEAST: Bayesian Evolutionary Analysis by Sampling Trees". *BMC Evolutionary Biology* 7: 214.
- Dumézil, Georges. 1965. *Le livre des heros: Legendes sur les Nartes*. Paris: Gallimard.
- Earl, Dent A., & Bridgett M. von Holdt. 2012. "STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno method". *Conservation Genetics Resources* 4(2): 359–361. Doi: 10.1007/s12686-011-9548-7.
- Eldredge, Niles, & Stephen Jay Gould. 1972. "Punctuated Equilibria: An Alternative to Phyletic Gradualism". In *Models in Paleobiology*. Ed. T.J.M. Schopf. San Francisco: Freeman Cooper. Pp. 82–115.
- Falush, D, M. Stephens, & J.K. Pritchard. 2003. "Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies". *Genetics* 164: 1567–1587.
- Frazer, James George. 1921. *Apollodorus: The Library II*. London: William Heinemann / New York: G.P. Putnam's Sons.
- Frobenius, Leo. 1996. *Contes Kabyles II: Le Monstrueux*. Trans. M. Fetta. Aix-en-Provence: Edisud.
- Frog. 2011. "Circum-Baltic Mythology? – The Strange Case of the Theft of the Thunder-Instrument (ATU 1148b)". *Archaeologia Baltica* 15: 78–98.
- Frog. 2013a. "The Parallax Approach: Situating Traditions in Long-Term Perspective". In *Approaching Methodology*. Ed. Frog & Pauliina Latvala with Helen F. Leslie. 2<sup>nd</sup> rev. edn. *Annales Academiae Scientiarum Fennicae Humaniora* 368. Helsinki: Academia Scientiarum Fennica. Pp. 101–131.
- Frog. 2013b. "Revisiting the Historical-Geographic Method(s)". In *Limited Sources, Boundless Possibilities: Textual Scholarship and the Challenges of Oral and Written Texts*. Ed. Karina Lukin, Frog & Sakari Katajamaki. RMN Newsletter 7. Helsinki: Folklore Studies, University of Helsinki. Pp. 18–34.
- Frog. 2014. "Germanic Traditions of the Theft of the Thunder-Instrument (ATU 1148b): An Approach to *Prymskviða* and Þórr's Adventure with Geirröðr in Circum-Baltic Perspective". In *New Focus on Retrospective Methods*. Ed. Eldar Heide & Karen Bek-Petersen. FF Communications 307. Helsinki: Academia Scientiarum Fennica. Pp. 120–162.
- van Gennep, A. 1909. *Religions, mœurs et légendes: Essais d'ethnographie et de linguistique (deuxième série)*. Paris: Mercure de France.
- Germain G. 1935. "Ulysse, le cyclope et les Berbères". *Revue de Littérature Comparée* 15: 573–623.
- Glenn, Justin. 1978. "The Polyphemus Myth: Its Origin and Interpretation". *Greece and Rome* 25(2): 141–155.
- Goddard, P.E. 1911. *Jicarilla Apache Texts*. Anthropological Papers of the American Museum of Natural History 8. New York: Trustees.
- Goddard, Pliny Earle. 1904. *Hupa Texts*. Berkeley: Berkeley University Press.
- Gould, Stephen Jay, & Niles Eldredge. 1977. "Punctuated Equilibria: The Tempo and Mode of Evolution Reconsidered". *Paleobiology* 3(2): 115–151.
- Greenhill, Simon J., Thomas E. Currie & Russell D. Gray. 2009. "Does Horizontal Transmission Invalidate Cultural Phylogenies?". *Proceedings of the Royal Society B: Biological Sciences* 276(1665): 2299–2306.
- Grimm, Wilhelm. 1857. *Die Sage von Polyphemus*. Berlin: Konigl. Akademie der Wissenschaften.
- Hackman, O. 1904. *Die Polyphem Sage in der Volksüberlieferung*. Helsinki: [Hackman].
- Hafstein, V.T.R. 2005. "Biological Metaphors in Folklore Theory: An Essay in the History of Ideas". In *Folklore: Critical Concepts in Literary and Cultural Studies*. Ed. Alan Dundes. London / New York: Routledge. Pp. 407–435.
- Hammer, Ø., D.A.T. Harper & P.D. Ryan. 2001. "PAST: Paleontological Statistics Software Package for Education and Data Analysis". *Palaeontologia Electronica*: 4(1). Available at: [http://palaeo-electronica.org/2001\\_1/past/issue1\\_01.htm](http://palaeo-electronica.org/2001_1/past/issue1_01.htm).
- Holland, B.R. et al. 2002. "δ Plots: A Tool for Analyzing Phylogenetic Distance Data". *Molecular Biology and Evolution* 19: 2051–2059.
- Huselsenbeck, J.P., & F. Ronquist. 2001. "MRBAYES: Bayesian Inference of Phylogeny". *Bioinformatics* 17: 754–755.
- d'Huy, J. 2012a. "Le Conte-Type de Polyphème". *Mythologie Française* 248: 47–59.
- d'Huy, J. 2012b. "Un ours dans les étoiles, recherche phylogénétique sur un mythe préhistorique". *Préhistoire du sud-ouest* 20(1): 91–106.
- d'Huy, J. 2012c. "Le motif de Pygmalion: Origine afrasienne et diffusion en Afrique". *Sahara* 23: 49–58.

- d'Huy, J. 2013a. "Polyphemus (Aa. Th. 1137): A Phylogenetic Reconstruction of a Prehistoric Tale". *Nouvelle Mythologie Comparée (New Comparative Mythology)* 1: 3–18.
- d'Huy J. 2013b. "A Phylogenetic Approach of Mythology and Its Archaeological Consequences". *Rock Art Research* 30(1): 115–118.
- d'Huy, J. 2013c. "A Cosmic Hunt in the Berber Sky: A Phylogenetic Reconstruction of Palaeolithic Mythology". *Les Cahiers de l'AARS* 16: 93–106.
- d'Huy, J. 2013d. "Les mythes évolueraient par punctuations". *Mythologie Française* 252: 8–12.
- d'Huy, J. 2013e. "Le motif du dragon serait paléolithique: Mythologie et archéologie". *Préhistoire du sud-ouest* 21(2): 195–215.
- d'Huy, J. 2013f. "Il y a plus de 2000 ans, le mythe de Pygmalion existait en Afrique du nord". *Préhistoires Méditerranéenne* 4. Available at: <http://pm.revues.org/814>.
- d'Huy, J. 2014a. "Mythologie et statistique. Reconstructions, évolution et origines paléolithiques du combat contre le dragon". *Mythologie française* 256: 17–23.
- d'Huy, J. 2014b. "Les contes peuvent-ils servir à remonter le temps". *Mythologie Française* 254: 14–16.
- d'Huy, J. 2014c. "Motifs and Folktales: A New Statistical Approach". *The RMN Newsletter* May, vol. 8: 13–29.
- d'Huy, J., & I. Dupanloup. 2015. "D'Afrique en Amérique: La bonne et la méchante fille (ATU 480)". *Nouvelle Mythologie Comparée / New Comparative Mythology* 2. Available at: <http://nouvellemythologiecomparee.hautetfort.com/archive/2015/02/19/julien-d-huy-et-isabelle-dupanloup-d-afrique-en-amerique%C2%A0la-5562363.html>.
- Karel, Jaronomir Erben. 1907. *Russian and Bulgarian Folk-lore Stories*. Trans. W.W. Strickland. London: G. Stranding.
- af Klintberg, Bengt. 2010. *The Types of the Swedish Folk Legend*. FF Communications 300. Helsinki: Academia Scientiarum Fennica.
- Krauss, Friedrich S. 1883. *Sagen und Märchen der Sudslaven in ihrem Verhältnis zu den Sagen und Märchen der übrigen indogermanischen Völkergruppen* I. Leipsic: W. Friedrich.
- Kroeber, A.L. 1907. "Gros Ventre Myths and Tales". *Anthropological Papers of the American Museum of Natural History* 1(3): 55–139.
- Krohn, Kaarle. 1926. *Die folkloristische Arbeitsmethode: Begründet von Julius Krohn und weitergeführt von nordischen Forschern*. Oslo: Aschehoug.
- Leroi-Gourhan, André. 1971. *Préhistoire de l'Art Occidental*. Paris: L. Mazenod.
- Lévi-Strauss, Claude. 1958. *Anthropologie structurale*. Paris: Plon.
- Lévi-Strauss, Claude. 1968. *L'origine des manières de table*. Paris: Plon.
- Lévi-Strauss, Claude et al. 2002. "Classification phylogénétique du vivant". *L'Homme* 162: 309–312.
- Lowie, Robert H. 1918. "Myths and Traditions of the Crow Indians". *Anthropological Papers of the American Museum of Natural History* 25(1): 1–308.
- Ludwig, Ross. 1863. *Erinnerungen und Mittheilungen aus Griechenland*. Berlin: R. Gaertner.
- Maddison, W.P. and D.R. Maddison. 2011. *Mesquite: a Modular System for Evolutionary Analysis*. Version 2.75. Available at: <http://mesquiteproject.org>.
- McAllister, J. Gilbert. 1949. "Kiowa-Apache Tales". In *The Sky Is My Tipi*. Ed. Mody C. Boatright. Publications of the Texas Folklore Society 22. Denton: University of North Texas Press. Pp. 1–141.
- de Nino, Antonio. 1883. *Usi e costumi abruzzesi* III. Florence: Tipographia di G. Barbera.
- Nunn, C.L., C. Arnold, L. Matthews, M. Borgerhoff Mulder 2010. "Simulating Trait Evolution for Cross-Cultural Comparison". *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365(1559): 3807–3819
- Oda, J. 2001. "Description of the Structure of the Folktale: Using a Multiple Alignment Program of Bioinformatics". *Senri Ethnological Studies* 55: 153–174.
- Opler, Morris Edward. 1938 [1969]. *Myths and Tales of the Jicarilla Apache Indians*. New York 1969: Kraus Reprinting.
- Opler, Morris Edward. 1940 [1969]. *Myths and Legends of the Lipan Apache Indians*. New York: Kraus Reprinting.
- Opler, Morris Edward. 1942 [1969]. *Myths and Tales of the Chiricahua Apache Indians*. New York 1969: Kraus Reprinting.
- Pagel, Mark et al. 2006. "Large Punctuational Contribution of Speciation to Evolutionary Divergence at the Molecular Level". *Science* 314(6): 119–121.
- Pala, M. et al. 2012. "Mitochondrial DNA Signals of Late Glacial Recolonization of Europe from Near Eastern Refugia". *American Journal of Human Genetics* 90: 915–924.
- Parsons E.C. 1929 [1969]. *Kiowa Tales*. New York: Kraus Reprinting.
- Patai, Raphael. 1998. *Arab Folktales from Palestine and Israel*. Detroit: Wayne State University Press.
- Peričić, P., et al. 2005. "High-Resolution Phylogenetic Analysis of Southeastern Europe Traces Major Episodes of Paternal Gene Flow among Slavic Populations". *Molecular Biology and Evolution* 22(10): 1964–1975.
- Poestion, J.C. 1886. *Laplandische Märchen, Volkssagen, Räthsel und Sprichwörter*. Vienna: C. Gerold.
- Pritchard, J.K., M. Stephens & P. Donnelly. 2000. "Inference of Population Structure Using Multilocus Genotype Data". *Genetics* 155: 945–959.
- Prym, Eugen, & Albert Socin. 1881. *Syrische Sagen and Maerchen aus dem Volksmunde*. Gottingen: Vandenhoeck & Ruprecht.

- Ralston, William Ralston Shedden. 1873. *Russian Folk-Tales*. London: Smith, Elder and Co.
- Rangel, T.F., et al. 2010. "SAM: A Comprehensive Application for Spatial Analysis in Macroecology". *Ecography* 33: 46–50.
- Richter, Fr von. 1889. "Litauische Märchen: Der einaugige Riese". *Zeitschrift für Volkskunde* 1: 87–93.
- Rohrich, L. 1976. *Sage und Märchen: Erzählforschung heute*. Freiburg: Herder.
- Ronquist, F., & J.P. Huelsenbeck. 2003. "MRBAYES 3: Bayesian Phylogenetic Mixed Models". *Bioinformatics* 19: 1572–1574.
- Ross, R.M., et al. 2013. "Population Structure and Cultural Geography of a Folktale in Europe". *Proceedings of the Royal Society B: Biological Sciences* 280: 1756.
- Simms, Stephen Chapman. 1903. "Traditions of the Crow". Field Columbian Museum, Publication 85. Anthropological Series 2.6. Chicago: Field Columbian Museum. Pp. 281–324.
- Spence, Lewis. 1914. *The Myths of the North American Indian*. New York: Thomas Y. Crowell Company.
- won Stephanowissel, Karadschitsch, with W. Tochter. 1854. *Volksmärchen der Serben*. Berlin.
- von Stier, Gaals Georg. 1857. *Ungarische Volksmärchen*. Pesth: Heckenast.
- von Sydow, Carl Wilhelm. 1948 [1965]. "Folktale Studies and Philology: Some Points of View". In *The Study of Folklore*. Ed. A. Dundes. Englewood Cliffs: Prentice Hall. Pp. 219–242.
- von Sydow, Carl Wilhelm. 1927. "Folksagor-forskningen". *Folkminnen och Folktankar* 14: 105–137.
- Tehrani Jamshid J. 2013. "The Phylogeny of Little Red Riding Hood". *PLoS ONE* 8.11: e78871.
- Thompson, Stith. 1961. *The Types of the Folktale: A Classification and Bibliography: Antti Aarne's Verzeichnis der Märchentypen Translated and Enlarged*. 2<sup>nd</sup> rev. edn. FF Communications 3. Helsinki: Academia Scientiarum Fennica.
- Tolley, Clive. 2012. "On the Trail of Þórr's Goats". In *Mythic Discourses: Finno-Ugrian Studies in Oral Tradition*. Ed. Frog, Anna-Leena Siikala & Eila Stepanova. *Studia Fennica Folkloristica* 20. Helsinki: Finnish Literature Society. Pp. 82–119.
- Uther, Hans-Jörg. 2004. *The Types of International Folktales: A Classification and Bibliography, Based on the System of Antti Aarne and Stith Thompson: Tales of the Stupid Ogre, Anecdotes and Jokes, and Formula Tables*. Helsinki: Suomalainen Tiedeakatemia.
- Venditti, Chris et al. 2006. "Detecting the Nodedensity Artifact in Phylogeny Reconstruction". *Systematic biology* 55(4): 637–643.
- Vialou, Denis. 1987. *L'Art des Cavernes: Les Sanctuaires de la Préhistoire*. Paris: Le Rocher.
- Vinson, Julien. 1883. *Le folk-lore du pays basque*. Les littératures populaires de toutes les nations 15. Paris: Maisonneuve.
- Wagner, Peter J., & Douglas H. Erwin. 1995. "Phylogenetic Patterns as Tests of Speciation Models". In *New Approaches to Studying Speciation in the Fossil Record*. Ed. D.H. Erwin & R.L. Anstey. New York: Columbia University Press. Pp. 87–122
- Webster, Andrea J., Robert Payne & Mark Pagel. 2003. "Molecular Phylogenies Link Rates of Evolution and Speciation". *Science* 301: 478.
- Webster Wentworth. 1879. *Basque Legends: Collected, Chiefly in the Labourland*. London: Griffith & Farran.
- Wichmann, S., et al. 2011. "Correlates of Reticulation in Linguistic Phylogenies". *Language Dynamics and Change* 1: 205–240.
- Wissler, Clark, & D.C. Duvall. 1908. *Mythology of the Blackfoot Indians*. Anthropological Papers of the American Museum of Natural History 2.1. New York: American Museum of Natural History.

### Internet Sites

Glottolog. <http://glottolog.org/>.

WALS = World Atlas of Language Structures. <http://wals.info/>.

## De situ linguarum fennicarum aetatis ferreae, Pars I

Frog and Janne Saarikivi, University of Helsinki

*Abstract: This article is the first part of a series that first employs a descendant historical reconstruction methodology to reverse-engineer areas where Finnic languages were spoken especially during the Iron Age (500 BC – AD 1150/1300). This opening article of the series presents a heuristic cartographic model of estimated locations of groups speaking Finnic languages and their neighbours in ca. AD 1000.*

The aim of this article is to provide the first of three maps of the Uralic-speaking peoples in Northwest Europe in three approximated periods: ca. AD 1000, AD 1 and a map indicating the linguistic *Urheimats* of reconstructed intermediate proto-languages within the Uralic language family (Proto-

Finnic, Proto-Sámi, Proto-Mordvin, etc.). For reasons of length, the aim of providing three different maps which involve different materials and present different issues has required presenting the investigation as a series. The present article is only the first part