Polyphemus (Aa. Th. 1137): A phylogenetic reconstruction of a prehistoric tale.

Julien D’Huy

To cite this version:


HAL Id: hal-00826004
https://hal.archives-ouvertes.fr/hal-00826004
Submitted on 28 Jan 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.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License
Polyphemus (Aa. Th. 1137)

A phylogenetic reconstruction of a prehistoric tale

Julien d’Huy,
Institut Marcel Mauss (UMR 8178) - Équipe LIAS - École des Hautes Études en Sciences Sociales (France)
dhuy.julien@yahoo.fr
http://ehess.academia.edu/JuliendHuy

Abstract: Mythology, like genetics and language, provides essentials elements in the understanding of human history; phylogenetic trees based on mythological versions can allow to reconstruct the history and prehistory of human cultures right back the Palaeolithic period. Indeed, mythological versions are highly conservative. We also can reconstruct the primitive version of a mythological family using phylogenetic reconstructions of ancestral state.

Keywords: mythological structure; mythological evolution; evolution; phylogenetic comparative method; networks; cultural phylogenetics; cultural evolution; ancestral state; vertical transmission; ancestral state; Polyphemus.

Résume: La mythologie, comme la génétique et les langues, fournit des éléments essentiels à la compréhension de l'histoire humaine. Des arbres phylogénétiques basés sur des versions mythologiques peuvent permettre de reconstituer l'histoire et la préhistoire de cultures humaines en remontant jusqu'à la période paléolithique. En effet, les versions mythologiques sont très conservatrices. Nous pouvons également reconstruire la version primitive d'une famille mythologique à l'aide de reconstructions phylogénétiques d'un état ancestral.

Mots clés : structure mythologique; évolution mythologique; évolution; méthode comparative et phylogénétique; réseaux; phylogénétique culturelle; évolution culturelle, état ancestral ; transmission verticale; état ancestral, Polyphème.

The Finnish school, which has an empirical and positivistic approach, tried to trace, collect and categorize all the variants of a story to reconstruct its history. It tried to establish the pure primeval tale (Urmärchen) from which
the versions are originated. Despite the initial enthusiasm, the reconstructive ambitions of the Finnish school have been strongly criticized. Criticisms of this method can be divided into four main categories: first, it gives a preponderant influence to oral tradition vs. literacy influence; second, the attempt to find the place of origin of tales seems to be doomed to failure, because the evidence has not been evenly collected among countries; third, the tale travelled independently of human migrations; fourth, there is no attempt to effect a reduction that would show how two or more seemingly different themes stand in a transformational relationship to each other.

Recent progresses relating to the development of computational pylogenetic tools, however, may help to resolve the main four problems faced by the Finnish school. First, phylogenetic methods analyze taxa as brothers or cousins, and not in parentage (Levi-Strauss 2002); we also don’t need to presume a gap between the true folktale and literary adaptations. Second, the problem of the origin can be solved by analyzing discrete characters to find the best possible tree. Third, the mythological evolution doesn’t need to follow the same way than the genetic evolution; moreover, computational methods can be used to verify that there is a real tree-like signal in the data. Fourth, there are at least two additional principles (variation and selection) of the folklore transmission being compatible with both evolutionist and structural treatments: the more two myths diverge or transform each other, the more distant is their genetic relationship, geographically and temporally (d'Huy 2012d).
There are many parallels between the process of biological and mythological evolution (d'Huy 2012a, b, d, 2013; see the table) and tools from evolutionary biology are being imported to analyze linguistic and cultural phenomena (Nunn 2011). Additionally, one of the key things about evolutionary science involve reconstructing ancestral states. We also can use phylogenetic comparative methods to make inferences about what the first version of a myth was like in the past and how it has changed over time. The main challenge in this research will be the lack of empirical data: prehistoric mythology leaves almost no traces.

Fig.1.: Myths, like biological species, evolve by a process of descent with modification. Many parallels can be drawn.

<table>
<thead>
<tr>
<th>GENETIC SYSTEMS</th>
<th>MYTHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete heritable units: e.g. the four nucleotides, codons, genes and individual phenotypes.</td>
<td>Discrete heritable units: e.g. the mythemes, motifs, tale-type.</td>
</tr>
<tr>
<td>Slow rate of evolution.</td>
<td>Fast or slow rate of evolution.</td>
</tr>
<tr>
<td>Parent-offspring, occasionally clonal.</td>
<td>Parent-offspring, generational transmission, teaching, writing (more recent).</td>
</tr>
<tr>
<td>Mutations (e.g. point mutations, insertions, deletions, loss of heterozygosity).</td>
<td>Innovations (e.g. recastes, variations, mistakes)</td>
</tr>
<tr>
<td>Natural selection of traits (individuals who possess certain variants of the trait survive and reproduce more than individuals who possess other variants).</td>
<td>Social selection of traits (e.g. societal trends, conformity, respect of tradition)</td>
</tr>
<tr>
<td>Allopatric or sympatric speciation.</td>
<td>Geographical or social separation.</td>
</tr>
</tbody>
</table>
Hybridization. | Mixture between two myths.
---|---
Horizontal transmission (e.g. hybridization, transposons and viruses). | Imposition or extralineal borrowing.
Geographic cline. | Mythological transformations.
Fossils. | Ancient myths that have survived in written form.
Extinction. | Myths death.

The reconstruction of Polyphemus tale is a textbook case. Many researchers have tried to reconstruct the protomyth, the significance which lies at the root of the story, albeit unsuccessful (for a review, see: Glenn 1978). The most complete attempt for reconstructing the first version of Polyphem was the Hackman's historical and geographical approach (Hackman 1904). Hackman supposed that the original form included I/ the blinding of the giant without his consent, II/ the flight of the hero under the belly of a ram, III/ the moment at which the hero gives a false name, IV/ the magic ring episode. According to Hackman, the versions would have travelled to northern Europe via Asia Minor or Greece. Yet Hackman's conclusions were based on statistical criteria that are unclear. Moreover, the author encountered difficulties to arrive at a precise definition of the “original version” of Polyphemus story because of the total lack of explanation for the criteria used to limit the number of versions included into the corpus (Calame 1995: 143). The problem of the physical, geographical origin of the story also seems insoluble (Glenn 1978).

We defined the Polyphemus type as a tale in which a person gets into a homestead – a master of animals' or a monstrous shepherd’s; the host can
kill the hero, but he escapes sticking to the hair of an animal who is going out or under an animal’s skin. We tried to solve the specific type of problems we were facing by using powerful statistical tools that are normally used in biology.

We first tested the hypothesis that mythemes, as genes, retained a phylogenetic signature with 79 mythemes in 24 versions of Polyphemus tales belonging to the European and North Amerindian areas (d'Huy 2012c).

We used a phylogenetic algorithm, *Bio Neighbor Joining* (implemented in the program *Splitstree* 3.2.) to explore global patterns of Polyphemus tales. We obtained a main delta-score of 0.2924. The 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); it also can be used to estimate the degree of conflict (Gray *et al.* 2010). The tree obtained agreed with a logical and historical reconstruction of the myth. Moreover, our results showed that the diversification of Polyphem versions was consistent with the distribution of the haplogroup X2. This haplogroup was present in the prehistoric Europe (Lacan *et al.* 2011, Deguilloux *et al.* 2011), and has moved from Beringia directly to the North American regions around 18,000 BC, before the Bering land bridge disappeared (Perego *et al.* 2010).

We decided to go further. We created a new and bigger database including many typological variations of the Polyphemus versions. The data used came from Ojibwa people (Désveaux 1988: 83), Atsina people (Kroeber 1907: 65-67), Niitsítapi people (Wissler et Duvall 1908: 50-52), Kootenays people (Boas 1918: 213-219, 303-304), Jicarilla Apache people (Goddard

A structural questionnaire of Polyphemus type was developed on the basis of the known variations in the studied versions. The questions focused on mythemes that are irreducible, unchanging elements in a story. Logically dependent features were eliminated. The mythemes were coded as presence (1) or absence (0) for each version to produce a binary matrix of 98
mythemes in 44 versions. The code incorporated data uncertainty (with the symbol ?).

Each mythology borrows myths from neighboring cultures in various contexts. Yet examining network of versions using NeighborNet with Splitstree3.2. reveals a moderate tree-like signal in the tree (fig.2). The resulting network reflects a real (but not dominant) conflicting signal found in structural traits (box-like structures). We must also stress the fact that versions are organized at a very high level with an Eurasian and an Amerindian cluster. Additionally, the graph correctly brings some of the versions together into well-known language or cultural families, with Basque people (Basque people Cerquand B, C, D / Basque people Webster), Gascon people (Gascon 1 and 2), Mediterranean people (first cluster : Greek 1, Berber 3, Serbian ; second cluster with a Berber / Italian group), Black Sea people (first cluster : Mingrelia / Ossetian ; second cluster : Abaza 1 / Oghuz turk), Apache people (Lipan apache and Jicarilla apache) and Algonquian people (Ojibwe / Niitsitapi / Kiowa 1 and 2 / Atsina). Mixed groups could be due to rapid population expansions, conflicting stronger signals (that may be produced by borrowings) and/or non-tree-like descent process (such as hybridization with another myth). Yet these data provide a strong argument in favour of a deep stability of myths.
Fig. 2: Splits graphs showing the results of NeighborNet analysis. The delta score was 0.3639 and the average Q-residual = 0.03148. The network shows the signal of grouping mythological versions.

The data was then analysed in the phylogenetic package *Mesquite 2.75* using a simple model to calculate the parsimony treelength of the tree and matrix (treelength: 334). Then the tree was rearranged by subtree pruning and regrafting. Finally, we rooted the tree between the Ojibwa and the Valais (fig.3).
According to the Sapir's age-area hypothesis, the area of greatest divergence in a linguistic family is said to point to the original homeland of the family; when we look at the tree, the greatest divergence is evidently between Ojibwa version and the rest of the family; consequently, the Ojibwa version is the most archaic Amerindian version. The Valais is probably the most archaic European version; it is intermediate between European and Amerindian corpus, with a master of the beasts similar to Amerindian versions. Additionally the closeness in our previous work between Ojibwa and Valais versions corroborates this hypothesis (see d’Huy 2012c).

To check the results, we used MrBayes 3.2.1. to infer the posterior distribution of phylogenetic tree for the 44 versions. We run an ordinary MCMC (Markov Chain Monte Carlo) analysis for 15.000.000 generations. The trees were sample every 1.000 generations. At the end of the run, the
average standard deviation of split frequencies was 0.0053. The fact that stationarity had been reached was controlled with Tracer 1.5.0. Summaries were based on a total of 22,502 samples from 2 runs. Each run produced 15,001 samples of which 11,251 samples were included. The tree was rooted with the Valais people version (fig.4).

Fig.4: Consensus tree from the Bayesian analysis of Polyphemus versions.
Node labels show posterior probability (the number is a ratio of one).
The Retention Index (RI) for the tree calculated with *Mesquite 2.75* was 0.612 for the Mesquite tree (fig.3), and 0.792 for the Bayesian tree (fig.4). The RI measures the degree to which each trait is shared by two or more taxa and their most recent common ancestor, whose own ancestor in turn does not possess the trait (synapomorphy). An RI close to 1 is the sign of a tree with a stable and slow evolution; conversally, an RI close to 0 is the sign of a tree with an unstable or a fast evolution. The both RI we obtain (0.612 / 0.792) are similar or greater than the mean RIs for the biological data sets presented by Lycett and *al.* (2009), whose main RI is 0.61. The Lycett's biological data sets was probably structured by speciation. So the horizontal transmission (from mother to daughter populations) should be the dominant evolutionary process in both biological and mythological data (see also d’Huy 2012d, e, 2013). Additionally, we obtained a delta-score of 0.3639 for the figure 2. The mean delta-score for European versions was 0.33, and the mean delta-score for Amerindian versions was 0.29. We can conclude that Amerindian mythology is more conservative than the European one.

The trees obtained with *Mesquite* (fig.3) and *MrBayes* (fig.4) are not exactly similar. Yet they also are partially consistent with our previous results; the results show a remarkably consistent pattern geographically speaking (North America / Europe; nearest geographical versions tended to form sister clades); the geographical area of the myth is consistent with the haplogroup X2 area in both North America and Europa; Swiss and Ojibwa versions conserve their most archaic place and their proximity can not be explained by a recent areal contact; the closeness between Gascons and
Berbers (d'Huy 2012c) is confirmed. On figure 3 and 4, red branches are located in northeastern North America; orange branches represent the southern region of United States of America; yellow branches represent the northwestern regions of United States. Cyan branches represent the Swiss region; blue branches represent the Mediterranean Sea area; magenta branches represent the Black Sea area; Sienna branches are located in northeastern Europe; green branches are located in the Pyrenean area; grey branches represent the north-western Europe.

According to Malhi and Smith (2002), haplogroup X may have had much more expansive distribution in North America and probably in Europa. The Great Lakes and Near East were probably only a refugium, and not the starting point of initial expansion. So a map that exactly superposed the current haplogroup X diffusion and the Polyphemus versions occurrences don't make sense. Yet haplogroup X is one of the bigger haplogroup among many Amerindian tribes who know the Polyphemus stories, the Algonquian peoples – Ojibwa (25.7%), Niitsitapi (100% but only one sample has been tested), Atsina (11.1%) -, the Kiowa (40%) and the neighboring tribes of Kootenays: the Nuu-Chah-Nulth (11%-13%) and the Yakama (5%) (Smith et al. 1999). Moreover, versions of tribes with haplogroup X are the first branching in the phylogenetic trees, so these versions probably appear at the beginning of the Amerindian expansion. Strong concentration of haplogroup X2 is also found in the Near East, the Caucasus and Mediterranean Europe; and currently the haplogroup X2 is less strongly present elsewhere in Europe, probably because the last maximum glacial. It is also found among
the Berber. Consequently, its expansion agreed with the studied versions area.

Phylogenetic approaches have already provided many insights into the origin and distribution of language and culturally transmitted objects (for a synthesis, see Nunn 2011). Yet, to the best of our knowledge, it is the first time that a geographical distribution pattern allows a phylogenetic reconstitution to go back in time to the Palaeolithic period (to take another mythological instance, see d’Huy 2012d, 2012e).

A historical interpretation of the European branches is problematic. In the first version, conserved in Switzerland, the monster is a master of beasts; this version is largely lost. It could happen during the Late Glacial Maximum when population density in Northern Eurasia decreased. The populations that migrated toward the South (Syria, Greece) probably preserved at least partially the primitive story. Then a new version where the monster was a shelter may be disseminated thanks to successive migrations from the Mediterranean area: with the first migration, the myth was transmitted to northern regions (Lithuanian, Hungarian, then Russian and Lapp); with the second migration, the myth was transmitted to the Gascony; with the third migration, the myth was transmitted to the Basque area, then in the Black Sea area and in the north-western Europe. Note that links between the Ancient Greek trading cities along the Black Sea coast and the ancestors of the Northwest Caucasian peoples are undoubtedly due to extensive borrowings and influences in the myths of both peoples (Colarusso 2002: 202), e.g. Abaza people 1 and Mingrelia.
The tree with branches proportional to lengths (fig. 4) shows correlation between total path length from the root of the tree to its tips and number of speciation events (“node”) (fig.5). In other words, where many new mythological versions (“nodes”) appear, there is more total mythological change (longer path lengths). In evolutionary biology, this link indicates a punctuated equilibrium effect (Webster and al. 2003). According to this theory, most species don't change much for the greater part of their geological history; this state is called *stasis*. When significant evolutionary change occurs, it is generally restricted to rare and very fast events of branching speciation. If correct, newly sister mythological versions tend to diverge rapidly, and are followed by extended periods of stability with little net evolutionary changes. This punctuational effect may thus reflect a human capacity to enhance both the group identity and the identification of individuals with this group (*we think that* versus *they think that*)\(^1\) or it may be due to a mythological founder effect (d'Huy 2012e). The rest of the time, mythological versions remains inchanged for long period of time, sometimes for thousands of years.

\(^1\) For instance, a story of the origin of fire can be told by an amerindian to offset another story by an indian of another tribe (Goddard 1904: 197).
To control the absence of the “node-density artifact” (a well-known artifact of phylogenetic reconstruction that may lead us to believe in a false punctuated equilibrium effect), we calculated the coefficient of determination ($R^2$). 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. The $R^2$ with a linear regression is bigger than the $R^2$ with a logarithmic regression. The tree also doesn't seem to show the curvilinear trend that characterizes the node-density artifact (Venditti and al. 2006).

**Fig.6: Coefficient of determination of the linear regression ($R^2$)**

<table>
<thead>
<tr>
<th></th>
<th>linear function</th>
<th>logarithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>European versions</td>
<td>0.74</td>
<td>0.66</td>
</tr>
<tr>
<td>North Amerindian versions</td>
<td>0.89</td>
<td>0.87</td>
</tr>
<tr>
<td>European and North Amerindian versions</td>
<td>0.6</td>
<td>0.57</td>
</tr>
</tbody>
</table>
We used two phylogenetic comparative methods (Maximum Likelihood with model Mk1 and Parcimony reconstructions) implemented in Mesquite 2.75 to examine the probable form of the first state of the Polyphemus family. We applied these phylogenetic reconstruction methods to each mythems of the family. Then we only selected highly confident reconstructed mythems with more than 50% probability with both methods. In the text, we underline mythems with more than 75% probability; the logical connective written in brackets remains just suggestions.

There are at least two monsters; they live in a tent and possess a herd of wild animals. **Animals are locked. The hero is a hunter. He enters in the homestead of the monster uninvited**, with the express purpose of stealing something, animals or treasure. [Then] **the entrance is blocked with a great stone or a locked door. A monster tries to kill the hero and checks the animals that go away. The hero escapes by hiding under the belly of an animal.**

This abstract is reconstructed as a simplified version of the *Urmärchen*. Of course, it is important to remember that the real protomyth probably was as rich in complexity as the versions upon which the reconstruction is based. This protomyth informs us about what Palaeolithic European speakers saying. For instance, it documents evidence for a belief in a master of beasts.

The phylogenetic models used in this paper allowed us to: make accurate inferences about human migration (our data independently corroborate a
common and prehistoric origin of Amerindian and European haplogroup X2) taking into account the probability; test the impact of mythological borrowings between neighbouring tribes and the part of similarities appearing by chance; show that mythology evolves in punctuational burst; reconstruct palaeolithic states of a mythological family with quantification of confidence.

The author thanks Jean Peyresblanques and Patrice Lajoye for their help.


Bladé J.-F., Contes de Gascogne, 1 : Contes épiques, Paris, Maisonneuve et Ch. Leclerc, 1886.


Karadschitsch Won Stephanowissel, with the collaboration of W. Tochter (1854). *Volksmärchen der Serben*, Berlin.


Lycett S.J., Collard M. et McGrew W.C. (2009), « Cladistic analyses of behavioural variation in wild *Pan troglodytes* : exploring the chimpanzee
culture hypothesis », *Journal of Human Evolution*, 57, pp.337-349.