



HAL
open science

Une nouvelle méthode, rapide et efficace, pour reconstruire les premières migrations de l'humanité.

Julien d'Huy

► To cite this version:

Julien d'Huy. Une nouvelle méthode, rapide et efficace, pour reconstruire les premières migrations de l'humanité.. *Mythologie française*, 2015, 259, pp.66-82. halshs-01295102

HAL Id: halshs-01295102

<https://shs.hal.science/halshs-01295102>

Submitted on 4 Apr 2016

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 4.0 International License



Mythologie Française

Société de Mythologie Française

Bulletin trimestriel n° 259

Juin 2015



Contes de mensonge

Le chamane Lanval

La Vierge, les marais et la mort

Les contes et les migrations
de l'humanité

Une nouvelle méthode, rapide et efficace, pour reconstruire les premières migrations de l'humanité

par Julien d'HUY

J'aimerais proposer ici une méthode permettant de mettre en relation la distribution mondiale de certains contes avec plusieurs grandes migrations de l'humanité. Il s'agit de s'appuyer sur la diffusion des contes-type à travers le monde et d'y appliquer des outils phylogénétiques (d'Huy 2015). J'espère ainsi mettre en évidence une origine possiblement paléolithique d'une partie au moins des contes que nous connaissons encore aujourd'hui, et proposer une nouvelle méthode, concourant à celle de la génétique, pour reconstruire les premières migrations de l'humanité.

On me reprochera de ressusciter de vieilles lunes romantiques, telles qu'elles ont été exprimées au XIX^e siècle, lunes qui supposent, de manière quelque peu arbitraire selon certains, une continuité entre les contes d'aujourd'hui et les contes d'hier. Suite aux abus de certains auteurs, Dietz-Rüdiger Moser a pu écrire que « la question sur l'âge des contes reste sans réponse dans l'état actuel de la recherche, et l'on ne peut y répondre qu'avec prudence » (« Die Frage nach dem Alter von Märchen läßt sich beim heutigen Forschungsstand nur mit Zurückhaltung beantworten ») (Moser 1977 : 407). Pourtant, certains contes semblent bel et bien d'origine paléolithique, comme le récit de Polyphème (ATU 1137 ; d'Huy 2012a, 2013a),

le conte de la Bonne et de la Méchante Fille (ATU 480 ; d'Huy 2014), celui de la femme-oiseau (d'Huy 2011a, b), le conte d'Eglé (Sergent 1999) ou encore celui de Jean de l'Ours (Sergent 2009), parmi d'autres. Le personnage du Trickster semble lui aussi paléolithique (Berezkin 2010, 2014). De plus, l'usage d'outils statistiques a permis de faire de réelles percées dans le domaine de la mythologie comparée, en révolutionnant l'approche, et laissant bon espoir d'apporter un nouvel éclairage au problème.

I. MATÉRIEL

Je me suis appuyé sur la partie « Animaux sauvages » du catalogue de contes-type établi par Uther (2004), lui-même basé sur l'œuvre d'Aarne (1910) et Thompson (1961). La domestication des animaux (partie « Domestic animals »), et la relation actuelle ou récente des hommes et des bêtes sauvages (parties « Wild animals and domestic animals » et « Wild animals and Humans »), biaiseraient en effet les recherches vers un passé moins récent, agricole, où la relation de l'homme et de l'animal a changé, et conduire à un artefact dans les calculs statistiques. Mais la vie sauvage, à travers sa

permanence, pourrait assurer la pérennité de certaines histoires, et les contes-types la concernant méritent que l'on s'y intéresse. Il s'agit là d'une hypothèse d'importance sur laquelle repose mon analyse. Par ailleurs, il faut noter que le choix des animaux principaux de chaque récit varie selon les régions, et que les contes-type établis par Uther privilégient par conséquent la stabilité de la structure sur le zoème.

On pourrait se demander si la nature éminemment modifiable d'un conte - selon les mélanges, les coupes, les rajouts des récitants - ne le rend pas difficilement saisissable par un formalisme mathématique. C'est oublier un peu vite la définition et le rôle d'un conte-type. Les contes-types constituent des structures narratives que l'on trouve répétée en plusieurs pays ou chez plusieurs ethnies (au minimum trois, chez Uther), et qui persistent malgré les variations orales. Il s'agit d'entités théoriques, incluant les principaux personnages d'une histoire et leurs actions, entités souvent définies de longue date par la pratique et la tradition folklorique.

J'ai ainsi créé une matrice dont les lignes correspondaient aux ethnies ou aires culturelles étudiées, et les colonnes aux divers contes-type possiblement présents dans ces aires. Les aires retenues sont les mêmes que celles d'Uther; elles sont de plusieurs ordres: ethniques, géographiques ou linguistiques. Afin d'échapper au biais de l'inventaire inégal des patrimoines folkloriques, seules les aires comptant plus de vingt contes-types ont été retenues (à l'exception du Pays Basque, qui ne compte que dix-neuf contes-types, mais qui a été retenu en tant qu'isolat linguistique). Le chiffre vingt a été choisi en raison de sa valeur élevée et parce qu'à partir de ce seuil, une augmentation du nombre de contes-type pris en compte par les calculs ne modifie pas substantiellement les résultats. Le catalogue d'Hans-Jörg Uther ne permet pas, pour des raisons évidentes d'uniformisation des critères, de dénombrer le nombre de versions présentes dans chaque aire géographique (Uther 2004 : 14). J'ai par ailleurs exclu des analyses les catégories « Roms », « Juifs » et « Hispano-Américain », pourtant présentes dans l'inventaire d'Uther, trop larges ou trop

difficiles à associer à une aire géographique donnée.

La question de la contamination des sources orales par des sources écrites s'est également posée. Certains contes analysés ont été très tôt notés, dans les *Ysopets*, les fables d'Ésope et les branches du *Roman de Renart*, par exemple, ce qui a pu avoir une incidence sur leur diffusion. Or ces contes ne sont pour la plupart pas apparus *ex nihilo*, et trouvent leur source dans des traditions populaires bien plus anciennes. Par ailleurs, le passage à l'écrit a pu s'accompagner d'une dissimulation « à distance » dans une aire déterminée, et on peut prédire, si l'influence de l'écrit prédominait celle de l'oral, à ce que nos calculs groupent certaines régions ensemble, comme les pays d'Europe occidentale et centrale ayant connu la diffusion d'épopées animalières centrées sur le personnage du renard.

J'ai codé par 1 la présence avérée d'un conte-type dans une région donnée, par 0 sa non-reconsension. Au final, 109 récits folkloriques ont été codés pour 48 aires.

II. APPLICATION DE LA METHODE DITE DU *bio neighbor joining*

Utilisant le logiciel *Splitstree4* (Huson et Bryant 2006), j'ai d'abord d'abord établi un arbre non enraciné en utilisant la méthode dite du *bio neighbor joining* sur la base des distances *uncorrect p'* (figure 1). En bioinformatique, cette méthode permet d'élaborer des arbres de données phylogénétiques. L'arbre ainsi obtenu n'est pas polarisé. J'ai obtenu le graphe 1. Son score est de 0.39. Ce score oscille entre 0 et 1; plus faible est sa valeur, plus les données analysées ont tendance à s'organiser sous la forme d'un arbre. Il existe donc ici un message phylogénétique, mais celui-ci est faible, probablement tempéré par de nombreux emprunts (diffusion aréale d'une grande partie du folklore) et des inventions spontanées. La faiblesse du score n'est guère surprenante quand nous prenons en compte la profondeur historique où nous nous situons, et la richesse de l'esprit humain. Cependant, sa valeur indique bel et bien un signal phylogénétique faible.

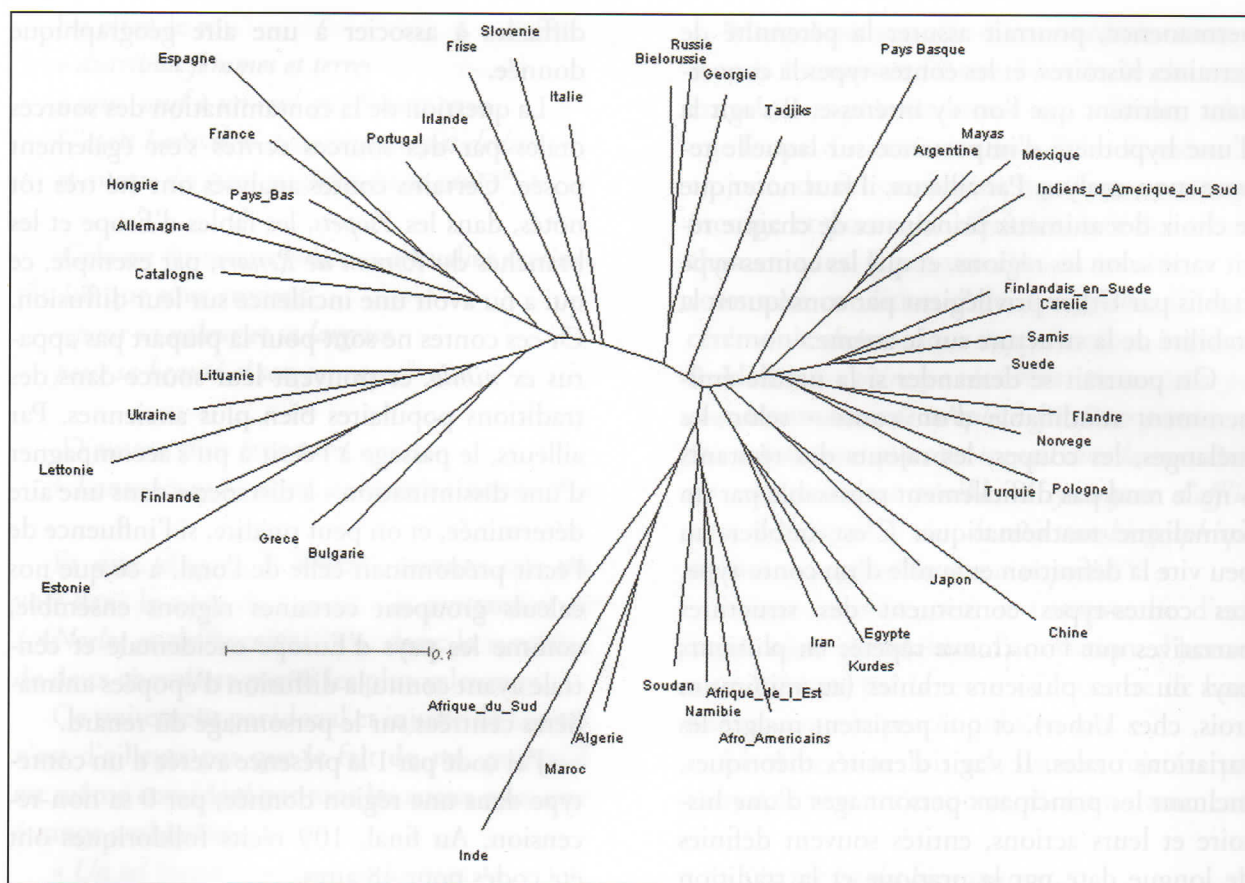


Figure 1 : arbre non enraciné construit en utilisant la méthode dite du bio neighbor joining sur la base des distances uncorrect ' p'.

Afin d'écarter l'hypothèse d'une diffusion essentiellement aréale (les traditions se transmettant de proche en proche), j'ai utilisé le logiciel SAM v.4.0 (Rangel et al. 2010) pour réaliser une matrice de distance Jaccard entre les différents folklores d'Eurasie. puis j'ai mis en regard le score obtenu entre chaque version et la distance géographique les séparant. Le positionnement géographique de chaque folklore a été déterminé par le point central de chaque pays (https://developers.google.com/public-data/docs/canonical/countries_csv) ou, le cas échéant, de chaque aire linguistique (<http://wals.info/languoid>). J'ai écarté les versions africaines et américaines car le logiciel SAM calcule le rapport entre le degré de divergence et la distance géographique en choisissant la distance la plus courte entre deux versions: l'Océan Pacifique et la mer Méditerranée, dont le franchissement ou le contournement se fait en des points particuliers, auraient ainsi rendu difficile, sinon impossible, l'intégration des ver-

sions non-eurasiatiques. Les résultats (figure 2) montrent que seulement 6,1 % de la variabilité propre à chaque folklore s'expliqueraient par la distance géographique, ce qui est légèrement inférieur aux résultats obtenus par Ross et al. (2013) pour le conte de la Bonne et de la Méchante fille (8,9 %). Cette faible corrélation me fait préférer une explication globalement généalogique à une explication diffusioniste pour rendre compte de la diffusion des folklores pris comme un tout. L'hypothèse d'une diffusion seulement aréale semble ici difficile à soutenir.

On observe, sur la figure 1, un premier ensemble, regroupant des nations africaines (Afrique du Sud, Maroc, Algérie, Soudan, Namibie, Afrique de l'Est, auquel s'ajoutent les Afro-Américains; seule véritable exception: l'Inde), connectées géographiquement à l'Égypte et à l'Asie occidentale (Iran, Kurdes), cette aire étant elle-même très proche de l'Extrême-Orient (Japon, Chine); la progression continue dans cette direction, incluant

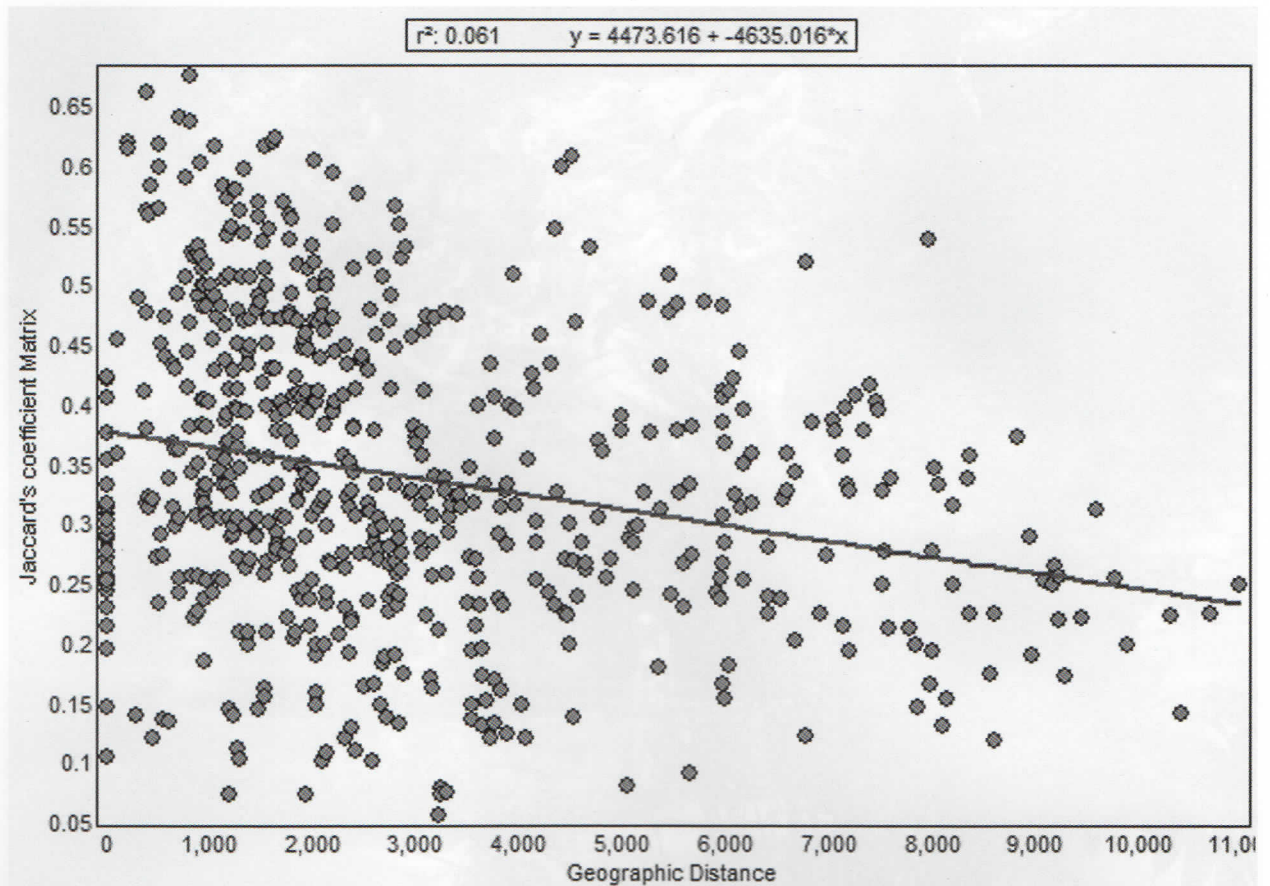


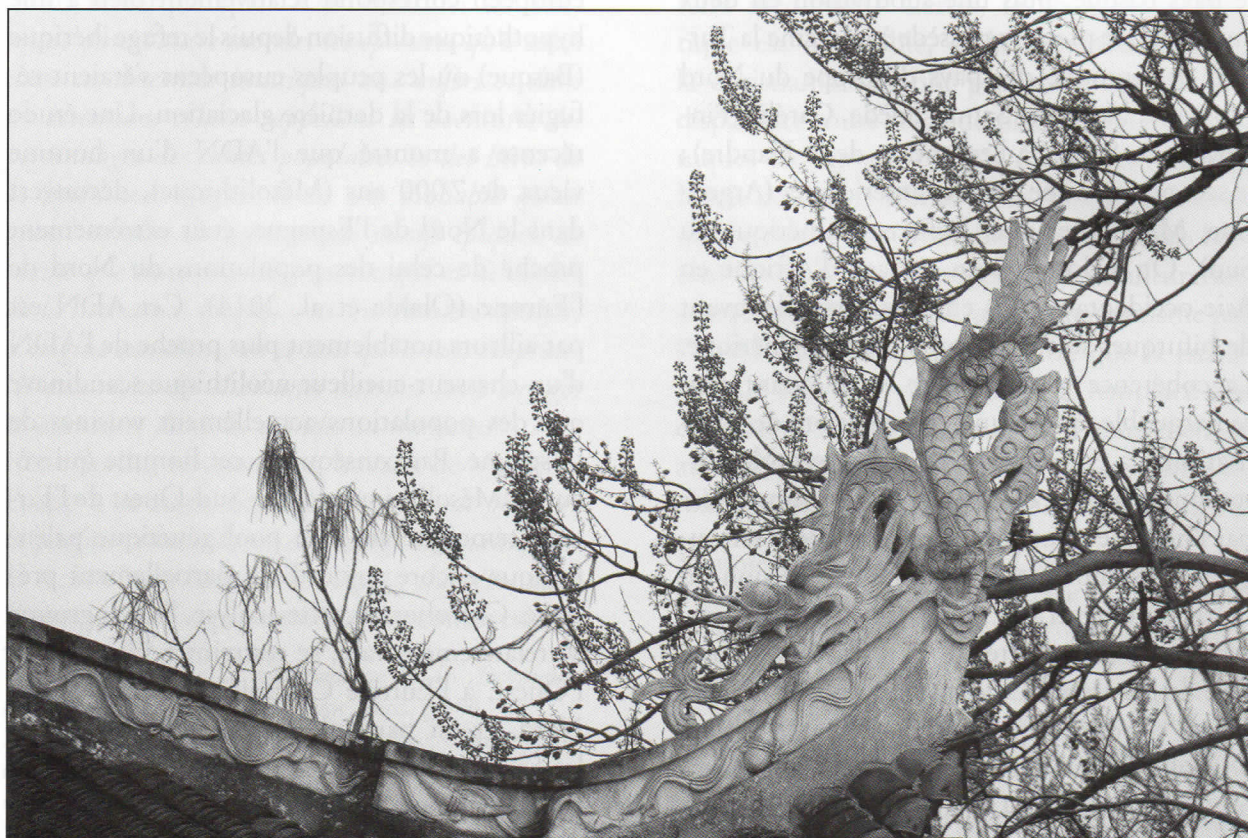
Figure 2: Test de Mantel, mettant en regard la distance structurelle séparant les versions du premier corpus et leur distance géographique.

le pays basque, puis une subdivision en deux ensemble : le premier possède à sa racine la Turquie et regroupe des pays d'Europe du Nord (Pologne, Norvège, Samis, Suède, Carélie, Finlandais en Suède, à l'exception de la Flandre) ; le second regroupe les aires américaines (Argentine, Maya, Mexique, Indiens d'Amérique du Sud). On a donc ici un passage d'Afrique en Asie occidentale, puis en Asie orientale, avant de bifurquer soit en Europe soit en Amérique. La cohérence géographique est d'autant plus remarquable qu'aucune information de localisation n'a été intégrée et qu'elle semble correspondre à l'une des grandes voies empruntées par l'homme pour conquérir le monde, retrouvée par exemple dans le cas de la diffusion du motif du dragon (d'Huy 2013b, 2014b). Corroborant l'ancienneté de la diffusion, le passage de l'Extrême-Orient (fig 2b) en Amérique n'a pu se réaliser que lorsque cela était encore possible, autrement dit, *a minima*, au Paléolithique supérieur. De même, le groupement

européen correspond relativement bien à une hypothétique diffusion depuis le refuge ibérique (Basque) où les peuples européens s'étaient réfugiés lors de la dernière glaciation. Une étude récente a montré que l'ADN d'un homme vieux de 7 000 ans (Mésolithique), découvert dans le Nord de l'Espagne, était extrêmement proche de celui des populations du Nord de l'Europe (Olalde et al. 2014). Cet ADN est par ailleurs notablement plus proche de l'ADN d'un chasseur-cueilleur néolithique scandinave que des populations actuellement voisines de l'Espagne. Par conséquent, cet homme qui vécut au Mésolithique dans le Sud-Ouest de l'Europe témoignerait d'un pool génétique paléolithique encore aujourd'hui partiellement préservé. Corroborant cette analyse, l'haplogroupe V, relativement rare, se restreint à l'Europe de l'Ouest, à l'Europe Centrale et à l'Europe du Nord, et est particulièrement fréquent parmi les Sames Skolt (52 %) et les Basques (12 %). Sa datation et sa distribution géographique



Fig 2b : Le dragon, un motif omniprésent en Chine (Quingyanzhen). Clichés B. Robreau.



correspondraient à la reconquête des territoires européens libérés des glaces, après la dernière grande glaciation (Torrioni *et al.* 2001). Notons enfin que des rapprochements entre certaines traditions orales basques et samies/nordiques ont déjà été réalisés, sur des bases statistiques (à propos du motif de la Chasse cosmique: d'Huy 2013c) ou plus empiriques (à propos de certains tabous portant sur l'ours: d'Huy 2013d), pouvant tous s'expliquer par une diffusion post-glaciaire depuis le refuge ibérique. Concluons en remarquant que l'inclusion dans ce groupe de la Turquie peut s'expliquer par l'existence d'un autre refuge lors de la dernière glaciation, au Proche-Orient (Cinnioglu *et al.* 2003 ; Pala *et al.* 2012 ; Peri *et al.* 2005).

Le groupe africain, très proche de l'Asie occidentale, l'est également de l'Asie centrale (Tadjik), région proche, sur le graphe et dans la réalité, du Caucase (Géorgie) à la base d'un ensemble regroupant Biélorussie et Russie. Une autre branche semble partir d'Asie centrale et, plaçant l'Italie et la Slovénie à la base du groupe, rassemble de nombreux pays européens ayant été pour la plupart sous domination romaine. On notera l'existence de trois sous-ensembles, le premier incluant la Finlande, l'Estonie, la Lettonie, l'Ukraine et la Lituanie; le second rapprochant la Catalogne, l'Allemagne, les Pays-Bas, la Hongrie, la France, l'Espagne et le Portugal; le dernier unissant la Grèce et la Bulgarie.

Ce que montre cette organisation des contes-type, et le parallèle que l'on peut en dresser avec les premières migrations de l'humanité, c'est que les contes portant sur les animaux sauvages seraient parmi les plus primitifs au monde. Ils se seraient diffusés en même temps que les premiers peuplements humains, avant de connaître, en certains lieux (par exemple en Europe) une réélaboration, due aux aléas de l'histoire humaine.

III. CALCUL DU NOMBRE OPTIMAL DE GROUPES PERMETTANT DE RENDRE COMPTE DE L'ENSEMBLE DU CORPUS

J'ai utilisé le logiciel *Structure* 2.3.4 (Pritchard *et al.*, 2000; Falush *et al.*, 2003)⁽¹⁾ pour corroborer la structure géographique des contes. L'algorithme permet normalement de reconstruire la structure d'une population en se basant sur un ensemble de gènes recombinés (autrement dit, des gènes hérités de plus d'un parent). Il crée ensuite un nombre K de groupes, auquel chaque échantillon analysé est rattaché, soit complètement, soit en partie (si tel est le cas, sa probabilité d'appartenir à chaque groupe est donnée). *Structure* a déjà été employé pour étudier des familles de langues (Reesink, Singer *et* Dunn 2009) et de mythes (d'Huy 2014c).

Afin de calculer le nombre de groupes constituant cet ensemble, j'ai employé la méthode proposée par Pritchard *et al.* (2000) et améliorée par Evanno, Regnaut *et* Goudet (2005). Partant du corpus, j'ai imposé un nombre K de groupes, à chaque fois différent (de 1 à 15), permettant de classer les différentes aires en un nombre donné de groupes, puis j'ai comparé le delta K de chaque classement, afin d'identifier le classement le plus probable. Chaque calcul a été refait dix fois. Les résultats sont présentés dans la figure 3 (construite en utilisant le programme de Earl *et* vonHold 2012) : plus le score d'une proposition de structuration ($K = 1, 2, 3, \text{etc.}$) est élevé, plus la probabilité qui lui est associée est importante. Il est ainsi facile de constater que les différentes aires se divisent, de manière optimale, en quatre groupes, globalement similaires à ceux obtenus en utilisant la méthode du *neighbor joining*. Le logiciel *Structure* permet également de calculer l'indice

1. Length of Burnin Period: 40000 ; Number of MCMC Reps after Burnin: 60000 ; use Admixture Model ; Allele Frequencies are Correlated among Pops ; Estimate the Probability of the Data Under the Model ; les contes-types manquant pour une région donnée ont été considérés comme des allèles récessifs.

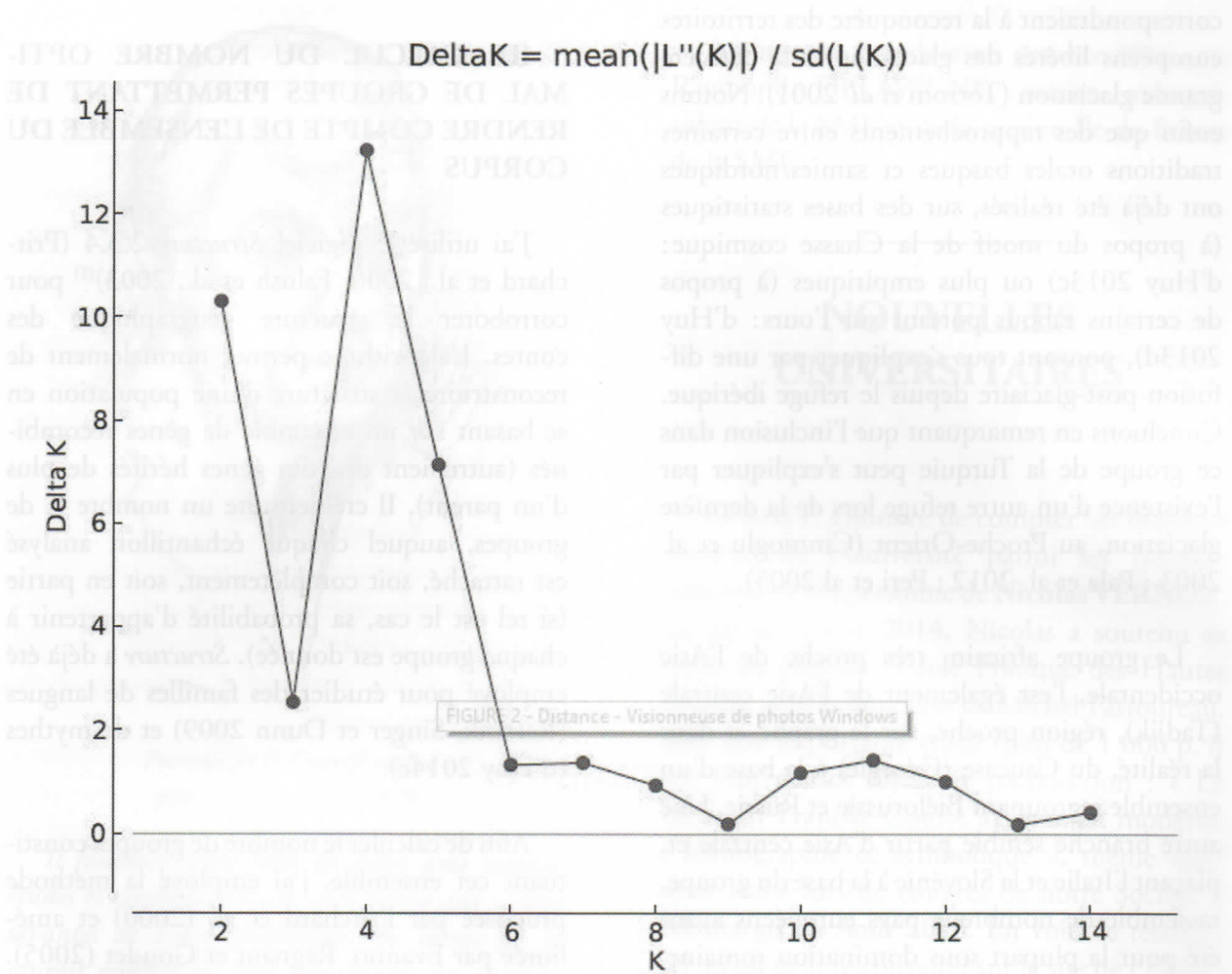


Figure 3 : Probabilité de chaque classement en k éléments pour le corpus sur les animaux sauvages.

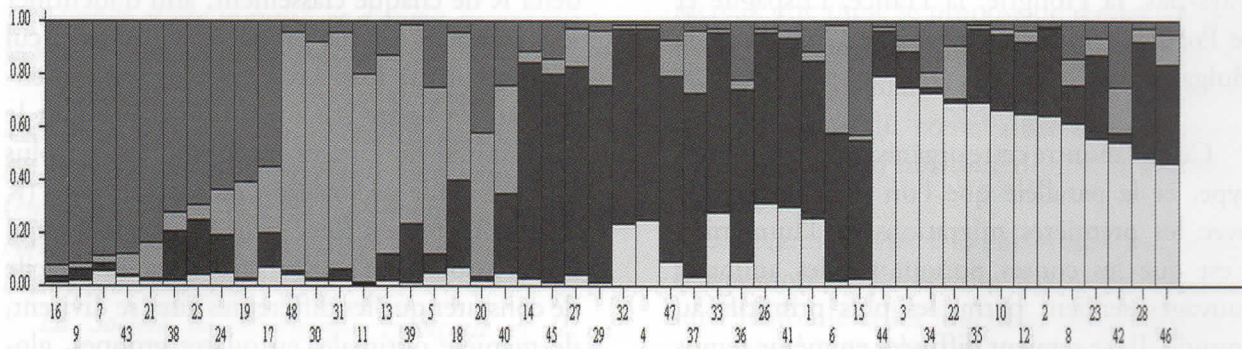


Figure 4: Part prise par quatre sources dans l'établissement de plusieurs traditions orales portant sur les animaux sauvages. 1. Pays basque ; 2. Afro-Américain ; 3. Algérie ; 4. Argentine ; 5. Bulgarie ; 6. Biélorussie ; 7. Catalogne ; 8. Chine ; 9. Pays-Bas ; 10. Afrique de l'Est ; 11. Estonie ; 12. Égypte ; 13. Finlande ; 14. Finlandais en Suède ; 15. Flandre ; 16. France ; 17. Frise ; 18. Géorgie ; 19. Allemagne ; 20. Grèce ; 21. Hongrie ; 22. Inde ; 23. Iran ; 24. Irlande ; 25. Italie ; 26. Japon ; 27. Carélie ; 28. Kurdes ; 29. Sami ; 30. Lettonie ; 31. Lituanie ; 32. Mayas ; 33. Mexique ; 34. Maroc ; 35. Namibie ; 36. Norvège ; 37. Pologne ; 38. Portugal ; 39. Russie ; 40. Slovénie ; 41. Indiens d'Amérique du Sud ; 42. Afrique du Sud ; 43. Espagne ; 44. Soudan ; 45. Suède ; 46. Tadjiks ; 47. Turquie ; 48. Ukraine.

de fixation (F_{st}), également appelé indice de différenciation, qui permet de mesurer la différenciation des populations à partir du polymorphisme génétique, en opposant la variabilité interne à chaque ensemble à la variabilité produite par l'écart entre les groupes. Un F_{st} de 0 indique une absence de variations entre deux populations, tandis qu'un F_{st} de 1 témoigne d'une complète différenciation.

- Le premier groupe rassemble les pays africains ou aires culturelles apparentées (Afrique de l'Est, Afrique du Sud, Afro-Américains, Algérie, Égypte, Maroc, Namibie, Soudan), ainsi que des pays d'Asie (Chine, Inde, Iran, Kurdes, Tadjiks). Son score F_{st} est de 0.3760.

- Le second groupe rassemble plusieurs pays ou populations d'Europe du Nord (Biélorussie, Carélie, Finlandais en Suède, Samis, Norvège, Pologne, Suède), les pays/aires culturelles américains du corpus (Argentine, Indiens d'Amérique du Sud, Maya, Mexique), auxquels s'ajoutent le Japon, la Turquie, le Pays basque et la Flandre. Son score F_{st} moyen est de 0.6164. Ces deux premiers ensembles correspondant probablement à la première expansion des contes paléolithique, que nous avons repérés.

- Le troisième ensemble réunit des pays d'Europe du Sud-Est (Bulgarie, Géorgie, Grèce), d'Europe du Nord (Estonie, Lettonie, Lituanie, Finlande) et d'Europe de l'Est (Ukraine, Russie). Son score F_{st} moyen est de 0.3683. Cet ensemble est géographiquement encerclé par les deux premiers groupes, laissant supposer une origine moins ancienne que ces derniers, mais encerclé à son tour le quatrième groupement.

- Ce dernier inclut la plupart des pays de l'Europe occidentale (Allemagne, Catalogne, France, Frise, Hongrie, Irlande, Italie, Pays-Bas, Portugal) ainsi qu'un pays d'Europe centrale (Hongrie). Le score F_{st} moyen de ce dernier groupe est de 0.3604. Le passage à l'écrit de certains contes et la circulation de ces textes sous la forme d'épopées animalières a probablement eu une influence sur la constitution de cet

ensemble, bien qu'il soit difficile de le démontrer (Uther 2006: 135-136).

Un polymorphisme d'un seul nucléotide (SNP, *single-nucleotide polymorphism*) est, en génétique, la variation d'une seule paire de bases du génome entre individu d'une même espèce. Une analyse de ces polymorphismes sur les chromosomes homologues et à l'échelle mondiale a montré que l'indice de fixation moyen entre les populations humaines autour du globe oscillait entre 0,052 et 0,130, ce qui indiquerait une variation de 5 à 13 % en moyenne entre le matériel autosomique de deux populations (Barbujani et Colonna 2011). Entre régions folkloriques, le F_{st} oscille entre 36,04 et 61,64 %, ce qui est bien supérieur à la variation génétique, et indique une grande fixité de ces aires qui, une fois établies, semblent peu échanger entre elles. Il serait donc bien plus facile de mélanger ses gènes à une population locale qu'à en modifier le folklore. Notons que les résultats corroborent ceux de Ross et *al.* (2013), dont les chiffres sont cependant plus faibles, et s'avèrent bien supérieurs aux F_{st} moyens basés sur des éléments culturels (0,08) et génétiques (0,0053) calculés par Bell et *al.* (2009).

On peut donc supposer une triple évolution : un substrat paléolithique, un substrat néolithique européen (?) et enfin un superstrat plus récent, proprement occidental. En utilisant une autre méthode statistique, j'obtiens donc des résultats extrêmement similaires à l'arbre de la figure 1.

On peut ici tenter une hypothèse hardie, inspirée de Roberts et *al.* 2014 : si l'homme de Néandertal parlait un langage moderne, et s'il avait existé des contacts culturels – après tout, une part de ses gènes coule dans nos veines – entre son espèce et la nôtre, si de plus Néandertal était capable de comportements symboliques présumant une mythologie, tels qu'enterrer ses morts (voir, pour une bonne synthèse, non exempte cependant de quelques approximations: Dediu et Levinson 2013), on pourrait supposer que le folklore ait été influencé par le folklore néandertalien. Si tel est le cas, et si les

		Delta score	Structure
1	Basque	0.38658	0.598
2	Atto-Américain	0.37426	0.647
3	Algérie	0.40447	0.752
4	Argentine	0.35236	0.716
5	Bulgarie	0.4058	0.625
6	Bébrussele	0.42147	0.556
7	Catalogne	0.38827	0.885
8	Chine	0.43627	0.620
9	Pays-Bas	0.38401	0.912
10	Afrique de l'Est	0.38091	0.668
11	Estonie	0.34951	0.785
12	Égypte	0.40123	0.651
13	Finlande	0.4089	0.748
14	Finlandais en Suède	0.38513	0.810
15	Flandre	0.4166	0.522
16	France	0.37516	0.918
17	Frise	0.39992	0.544
18	Géorgie	0.42642	0.559
19	Allemagne	0.36223	0.603
20	Grèce	0.38811	0.541
21	Hongrie	0.35841	0.830
22	Inde	0.43623	0.699
23	Iran	0.41239	0.565
24	Irlande	0.409	0.631
25	Italie	0.39666	0.690
26	Japon	0.386	0.643
27	Carélie	0.38826	0.787
28	Kurdes	0.39644	0.486
29	Samis	0.39157	0.743
30	Lettonie	0.36335	0.890
31	Lituanie	0.3929	0.890
32	Mayas	0.35022	0.738
33	Mexique	0.35609	0.680
34	Maroc	0.41737	0.730
35	Namibie	0.38049	0.694
36	Norvège	0.40377	0.648
37	Pologne	0.40049	0.694
38	Portugal	0.41602	0.720
39	Russie	0.41919	0.745
40	Sibérienne	0.41858	0.410
41	indiens d'Amérique du Sud	0.37854	0.640
42	Afrique du Sud	0.41822	0.546
43	Espagne	0.34922	0.873
44	Soudan	0.39017	0.796
45	Suède	0.39684	0.788
46	Tadjiks	0.41399	0.469
47	Turquie	0.40187	0.706
48	Ukraine	0.38155	0.900

Figure 5 : Score delta et part prise par la principale source à l'intérieur d'un folklore donné.

aires folkloriques ont bien la stabilité que les calculs leur prêtent, il devrait rester une trace de cette influence dans nos folklores contemporains. Aussi doit-on se demander s'il n'existerait pas une différence entre les aires folkloriques actuelles ayant été (hors de l'Afrique) ou n'ayant pas été (en Afrique) en contact avec Néandertal ? Deux indices semblent l'indiquer : 1/ l'Afrique constitue un groupe seul dans l'analyse par *clusters* réalisée grâce au logiciel *Structure* ; 2/ le groupe correspondant à la sortie d'Afrique possède le Fst le plus élevé de tous, ce qui pourrait s'expliquer par la rencontre avec une ou plusieurs autres espèces humaines. Ces résultats peuvent admettre d'autres explications

qu'une rencontre avec Néandertal, par exemple l'établissement d'une mythologie mondiale en deux étapes, la première suivant les premières migrations de l'homme, de sa sortie d'Afrique à l'Australie, la seconde apparaissant plus tard en Eurasie, surplantant partiellement la première et atteignant les Amériques (Witzel 2012). Les perspectives restent cependant vertigineuses et méritent sans aucun doute de futurs approfondissements.

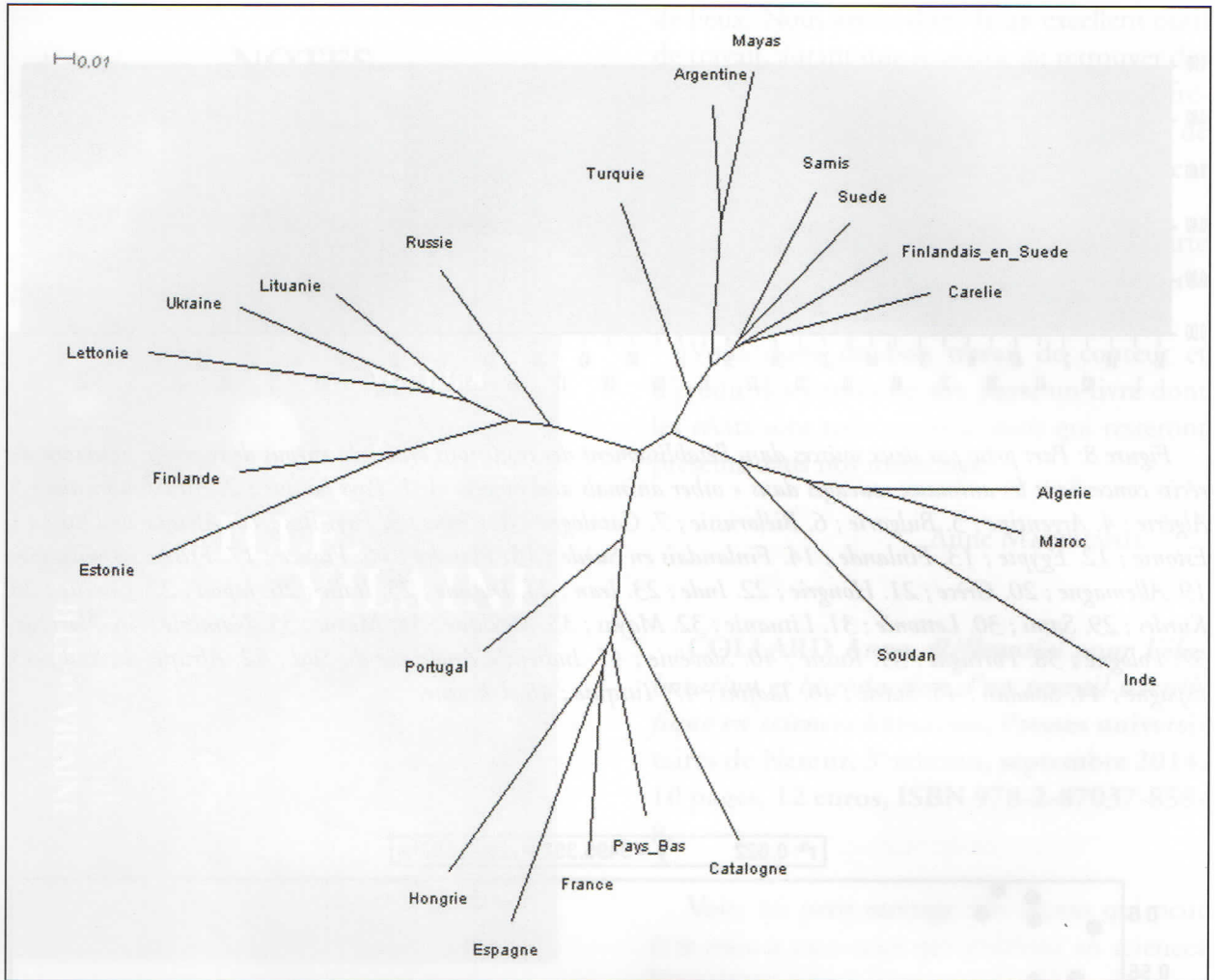


Figure 6 : reprise de l'arbre non enraciné construit en utilisant la méthode dite du bio neighbor joining, réalisé en prenant pour base les folklores relevant à plus de 70 % d'une même origine.

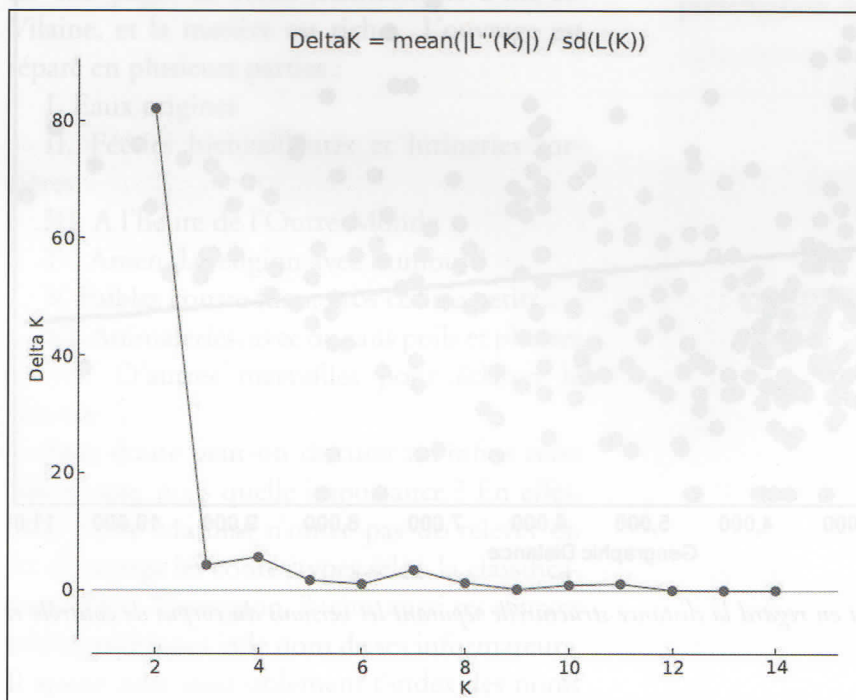


Figure 7 : Probabilité de chaque classement en k éléments pour le corpus de contrôle.

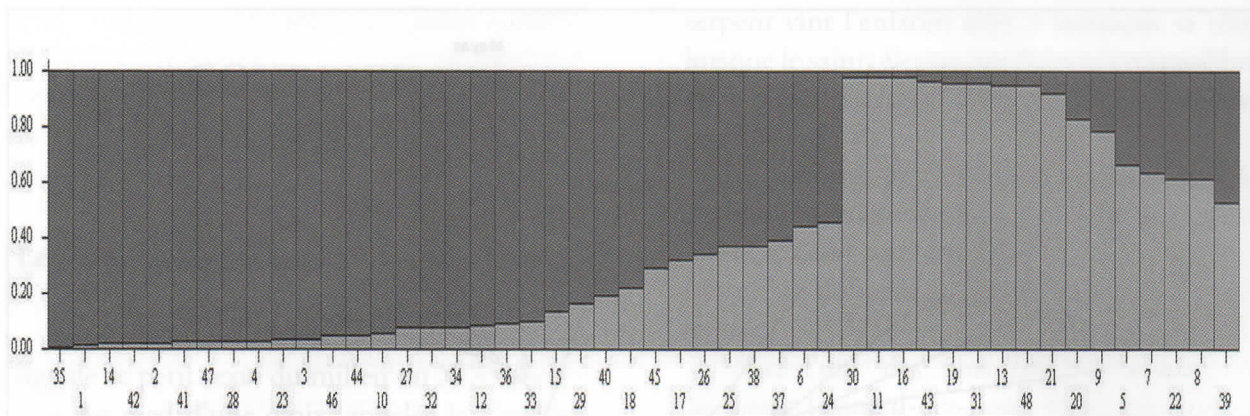


Figure 8: Part prise par deux sources dans l'établissement des traditions orales du corpus de contrôle (sélection des récits concernant les animaux sauvages dans « other animals and objects »). 1. Pays basque ; 2. Afro-Américain ; 3. Algérie ; 4. Argentine ; 5. Bulgarie ; 6. Biélorussie ; 7. Catalogne ; 8. Chine ; 9. Pays-Bas ; 10. Afrique de l'Est ; 11. Estonie ; 12. Égypte ; 13. Finlande ; 14. Finlandais en Suède ; 15. Flandre ; 16. France ; 17. Frise ; 18. Géorgie ; 19. Allemagne ; 20. Grèce ; 21. Hongrie ; 22. Inde ; 23. Iran ; 24. Irlande ; 25. Italie ; 26. Japon ; 27. Carélie ; 28. Kurdes ; 29. Sami ; 30. Lettonie ; 31. Lituanie ; 32. Mayas ; 33. Mexique ; 34. Maroc ; 35. Namibie ; 36. Norvège ; 37. Pologne ; 38. Portugal ; 39. Russie ; 40. Slovénie ; 41. Indiens d'Amérique du Sud ; 42. Afrique du Sud ; 43. Espagne ; 44. Soudan ; 45. Suède ; 46. Tadjiks ; 47. Turquie ; 48. Ukraine.

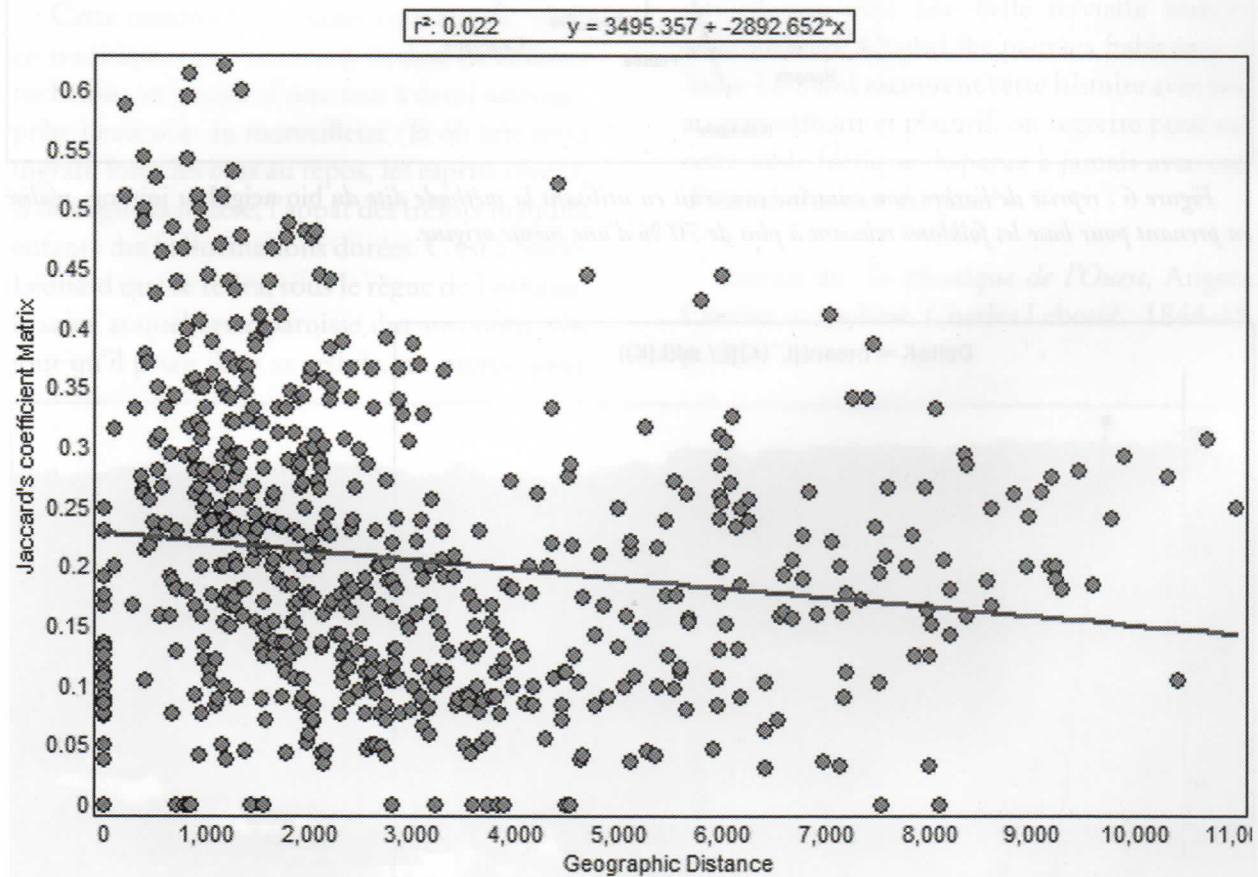


Figure 9: Test de Mantel, mettant en regard la distance structurelle séparant les versions du corpus de contrôle et leur distance géographique.

IV. RÉÉVALUATION DU GRAPHE PRÉCÉDEMMENT OBTENU

L'analyse de chaque folklore régional à l'aide du logiciel *Structure* permet de révéler les similarités entre régions. Dans la figure 4, les couleurs correspondent aux quatre groupes identifiés. Ici, aucune région n'a de couleur, donc d'origine, unique.

Suivant l'exemple de Bowerl (2012), j'ai exclu de l'analyse les régions folkloriques qui ne formaient pas un ensemble suffisamment uniforme (l'origine principale des folklores régionaux retenus devaient être égale à ou dépasser 70 %; l'Inde et le Soudan ont cependant été retenus car s'approchant à l'extrême de cette limite). Remarquons que la part prise par la source principale dans chaque folklore est en partie corrélée au score delta individuel de cet ensemble (Pearson : -0,46238 ; $p = 0,00093719$; Spearman : -0,44254 ; $p = 0,001635$), ce qui corrobore la capacité du score à prédire la plus ou moins bonne intégration d'un taxon (version d'un mythe / folklore) dans une phylogénie proposée (figure 5 ; voir d'Huy 2012b pour une première formulation de cette hypothèse). En me basant sur ce nouveau corpus, j'ai ensuite reconstruit un arbre grâce à la méthode du *bio neighbor joining* sur la base des distances *uncorrected p*. Le résultat obtenu (figure 6 ; score = 0,36) est semblable, dans son organisation, à mon premier essai, maintenant les principaux groupements et corroborant les résultats.

Cette méthode semble prometteuse : son application à d'autres corpus portant sur des familles de mythes / contes a permis de faire disparaître la majorité des incohérences constatées, qui consistait, dans la plupart des cas, à un mauvais placement des versions.

V. UTILISATION D'UN CORPUS-TEST

Comment contrôler ces résultats avec davantage encore de sévérité ? Il faudrait tenter de les répliquer à partir d'un nouveau corpus.

J'ai ainsi sélectionné, dans la partie « other animals and objects » de l'ouvrage d'Uther, l'ensemble des contes ne concernant que des animaux sauvages, excluant donc les interactions avec des animaux domestiques ou des hommes. J'ai obtenu un corpus de 73 contes-types (ATU 220, 220A, 221, 221A, 221B, 222, 222A, 222B, 223, 224, 225, 225A, 226, 227, 227*, 228, 229, 230, 231, 231*, 231**, 232C*, 232D*, 234, 234A*, 235, 235C, 236, 236*, 238, 239, 240, 240A*, 241, 242, 243*, 244, 244A*, 244B*, 244C, 246, 247, 248, 248A, 250, 250A, 253, 275A, 275B, 275C, 275C*, 276, 277, 277A, 278, 278A, 278A*, 279*, 280, 280A, 281, 283, 283B*, 283D*, 283H*, 288*, 288C*, 289, 291, 293, 293B*, 297B, 299) que j'ai codé selon leur absence ou présence pour les 48 pays / aires culturelles déjà étudiées. Les résultats obtenus seront nécessairement moins précis – le nombre de contes-type par aire géographique / culturelle pouvant être très inférieur à vingt – mais resteront de précieux indicateurs pour tester les résultats précédents.

Le logiciel *Structure* a été appliqué à ce nouveau corpus dans les mêmes termes que précédemment, montrant que ce dernier se divise de manière optimale en deux ensembles (figure 7). Le premier reprend les deux premiers groupes déjà identifiés, à l'exception de la Chine et de l'Inde qui sont associés au second ensemble. Le second groupe reprend l'essentiel du troisième et quatrième ensembles, à l'exception de la Géorgie, de la Frise, de l'Irlande, de l'Italie et du Portugal qui sont, quant à eux, rattachés au premier groupe. Ces résultats sont d'autant plus remarquables que le nombre de contes-type pris en compte est ici, comme je l'ai fait remarquer plus haut, bien inférieur à ce qui a été pris en compte lors de ma première analyse, et que le nombre minimal de vingt contes-type retenus précédemment pour une région est loin d'être respecté (le nombre de contes-type oscillant de 3 à 42 récits selon les régions étudiées).

Remarquons que le *Fst* moyen du premier groupe (0,7198), supposé paléolithique et influencé par des rencontres avec d'autres (ou une autre) espèces humaines (Neandertal,

Dénisovien, Homme de Florès, etc.), est une nouvelle fois largement supérieur au Fst du second groupe (0,0075). De plus, la part prise par la distance dans la variabilité des folklores eurasiatiques reste très faible, de l'ordre de 2,2%. Ces chiffres, qui doivent être pris avec précaution étant donné la faiblesse du corpus étudié, tendent cependant à corroborer les analyses précédentes.

VI. CONCLUSION

J'ai donc mis en évidence l'existence de deux grands ensembles de contes d'animaux, dont l'un est probablement paléolithique, en m'appuyant sur le chapitre « Animaux sauvages » de l'ouvrage d'Uther. Ces résultats ont été ensuite en grande partie répliqués en changeant l'approche statistique, puis à partir d'un autre ensemble de contes-type. Ils semblent donc relativement fiables, prouvant statistiquement que les contes perdurent dans le temps.

Il est par ailleurs intéressant de noter l'unité du groupe africain et du groupe américain, dont le peuplement a pourtant connu une histoire extrêmement complexe. De la même façon que les nouveaux arrivants dans une région ne parviennent pas à imposer entièrement leurs gènes, laissant perdurer le fond génétique des anciens habitants, les migrants adopteraient souvent en grand nombre les contes des peuples autochtones, ne parvenant pas à les supplanter par le nombre. Les contes se continueraient donc par « diffusion » verticale d'une vague de peuplement à l'autre, ce qui corroborerait l'hypothèse d'un substrat folklorique propre à une aire géographique donnée⁽²⁾ et permettrait de rendre compte d'une possible influence néandertalienne. Seule une pulsion migratoire suffisamment importante, ou un radical changement ontologique, pourrait provoquer l'émergence d'une nouvelle constellation

2. Un cas particulier est cependant celui des afro-américains, dont le graphe et les divers calculs montrent qu'ils se rattachent à l'Afrique. Voir également d'Huy et Dupanloup 2015.

folklorique. Si mon hypothèse s'avère correcte, il serait alors possible, en étudiant statistiquement la diffusion de certains motifs, de reconstituer les diverses vagues de peuplement ayant progressivement recouvert notre planète.

L'auteur remercie Hans-Jörg Uther, Patrice Lajoie, Jean-Loïc Le Quellec et Guillaume Oudaer pour les critiques apportées à une première version de ce texte.

BIBLIOGRAPHIE

AARNE, Antti. *Verzeichnis der Märchentypen, mit Hilfe von Fachgenossen*. Helsinki, FFC 3, 1910.

BARBUJANI, Guido et COLONNA, Vincenza. Genetic basis of human biodiversity: an update. In: *Biodiversity hotspots: distribution and protection of conservation priority areas* (éd F.E. Zachos, J.C. Habel), Berlin, Germany, Springer, 2011, pp. 97-119.

BELL, Adrian V., RICHERSON, Peter J., et MCELREATH, Richard. Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proceedings of the National Academy of Sciences*, 2009, vol. 106, no 42, p. 17671-17674.

BEREZKIN, Yuri. Tricksters Trot to America: Areal Distribution of Folklore Motifs. *Folklore: Electronic Journal of Folklore*, 2010, n°46, p. 125-142.

BEREZKIN, Yuri. World distribution of zoomorphic protagonists of the trickster tales. *Nouvelle Mythologie Comparée*, 2014, n°2, en ligne.

BOWERN, Claire. The riddle of Tasmanian languages. *Proceedings of the Royal Society B: Biological Sciences*, 2012, vol. 279, n°1747, p. 4590-4595.

CINNIOLU, Cengiz, KING, Roy, KIVISILD, Toomas, et al. Excavating Y-chromosome haplotype strata in Anatolia. *Human genetics*, 2004, vol. 114, n°2, p. 127-148.

DEDIU, Dan et LEVINSON, Stephen C. On the antiquity of language: the reinterpretation

tation of Neandertal linguistic capacities and its consequences. *Frontiers in Psychology*, 2013, vol. 4.

EVANNO, Guillaume, REGNAUT, Sebastien, et GOUDET, Jérôme. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, 2005, vol. 14, n°8, p. 2611-2620.

EARL, Dent A., et VonHOLDT, Bridgett M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, 2012, vol. 4, n°2, p. 359-361.

FALUSH, Daniel, STEPHENS, Matthew, et PRITCHARD, Jonathan K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 2003, vol. 164, n°4, p. 1567-1587.

HUSON, Daniel H. et BRYANT, David. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*, 2006, vol. 23, n°2, p. 254-267.

d'HUY, Julien. Le motif de la femme-bison Essai d'interprétation d'un mythe préhistorique (1ère partie). *Mythologie française*, 2011a, n°242, p. 44-55.

d'HUY, Julien. Le motif de la femme-bison Essai d'interprétation d'un mythe préhistorique (2e partie). *Mythologie française*, 2011b, n°243, p. 23-41.

d'HUY, Julien. Le conte-type de Polyphème: essai de reconstitution phylogénétique. *Mythologie française*, 2012a, n°248, p. 47-59.

d'HUY, Julien. Un ours dans les étoiles, recherche phylogénétique sur un mythe préhistorique. *Préhistoire du sud-ouest*, 2012b, vol. 20, n°1, p. 91-106.

d'HUY, Julien. Polyphemus (Aa. Th. 1137): A phylogenetic reconstruction of a prehistoric tale. *Nouvelle Mythologie Comparée*, 2013a, n°1, en ligne.

d'HUY, Julien. Le motif du dragon serait paléolithique: mythologie et archéologie. *Préhistoire du Sud-Ouest*, 2013b, vol. 21, n°2, p. 195-215.

d'HUY, Julien. A Cosmic Hunt in the Berber sky:: a phylogenetic reconstruction of Palaeolithic mythology. *Les Cahiers de l'AARS*, 2013c, n°15, p. 93-106.

d'HUY, Julien. L'Oreille de l'Ours. *Mythologie française*, 2013d, n°250, p. 10-14.

d'HUY, Julien. Les contes peuvent-ils servir à remonter le temps?. *Mythologie française*, 2014a, n°254, p. 14-16.

d'HUY, Julien. Mythologie et statistique: reconstructions, évolution et origines paléolithiques du combat contre le dragon. *Mythologie française*, 2014b, n°256, p. 17-23.

d'HUY, Julien. Une méthode simple pour reconstruire une mythologie préhistorique (à propos de serpents mythiques sahariens). *Les Cahiers de l'AARS*, 2014c, n°17, p. 95-104.

d'HUY, Julien. Phylogenetic tools and myths: reconstructing human prehistory. *The Journal of Brief Ideas*, 14 Février 2015, 10.5281/zenodo.15118

d'HUY, Julien et DUPANLOUP, Isabelle. D'Afrique en Amérique : la bonne et la méchante fille (ATU 480). *Nouvelle Mythologie Comparée*, 2015, n°2, en ligne.

MOSER, Dietz-Rüdiger. Altersbestimmung des Märchens. In : *Enzyklopädie des Märchens : Handwörterbuch zur historischen und vergleichenden Erzählforschung* (éd. Kurt Ranke, avec Hermann Bausinger, Wolfgang Brückner, Daniel Drascek... et al.), Berlin / New York, W. de Gruyter, tome 1, 1977, pp. 407-419.

OLALDE, Iñigo, ALLENTOFT, Morten E., SÁNCHEZ-QUINTO, Federico, et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature*, 2014, vol. 507, n°7491, p. 225-228.

PALA, Maria, ACHILLI, Alessandro, OLIVIERI, Anna, et al. Mitochondrial haplogroup U5b3: a distant echo of the epipaleolithic in Italy and the legacy of the early Sardinians. *The American Journal of Human Genetics*, 2009, vol. 84, n°6, p. 814-821.

PERI, Marijana, LAUC, Lovorka Bara, KLARI, Irena Martinovi, et al. High-resolution phylogenetic analysis of southeastern

Europe traces major episodes of paternal gene flow among Slavic populations. *Molecular biology and evolution*, 2005, vol. 22, n°10, p. 1964-1975.

PRITCHARD, Jonathan K., STEPHENS, Matthew, et DONNELLY, Peter. Inference of population structure using multilocus genotype data. *Genetics*, 2000, vol. 155, n°2, p. 945-959.

RANGEL, Thiago F., DINIZ FILHO, Jose Alexandre F., et BINI, Luis Mauricio. SAM: a comprehensive application for spatial analysis in macroecology. *Ecography*, 2010, vol. 33, n°1, p. 46-50.

REESINK, Ger, SINGER, Ruth, et DUNN, Michael. Explaining the linguistic diversity of Sahul using population models. *PLoS biology*, 2009, vol. 7, n°11.

ROBERTS, Sean G., DEDIU, Dan, et LEVINSON, Stephen C. Detecting differences between the languages of Neandertals and modern humans. In: *The Evolution of Language: Proceedings of the 10th International Conference* (éd. E. A. Cartmill, S. G. Roberts, H. Lyn, H. Cornish), Singapore, World Scientific, pp. 521-502.

ROSS, Robert M., GREENHILL, Simon J., et ATKINSON, Quentin D. Population structure and cultural geography of a folktale

in Europe. *Proceedings of the Royal Society B: Biological Sciences*, 2013, vol. 280, no 1756.

SERGEANT, Bernard. Un mythe lithuano-amérindien. *Dialogues d'histoire ancienne*, 1999, vol. 25, no 2, p. 9-39.

SERGEANT, Bernard. *Jean de l'Ours, Gargantua et le Dénicheur d'Oiseau*. La Bégude de Mazenc: Arma Artis, 2009.

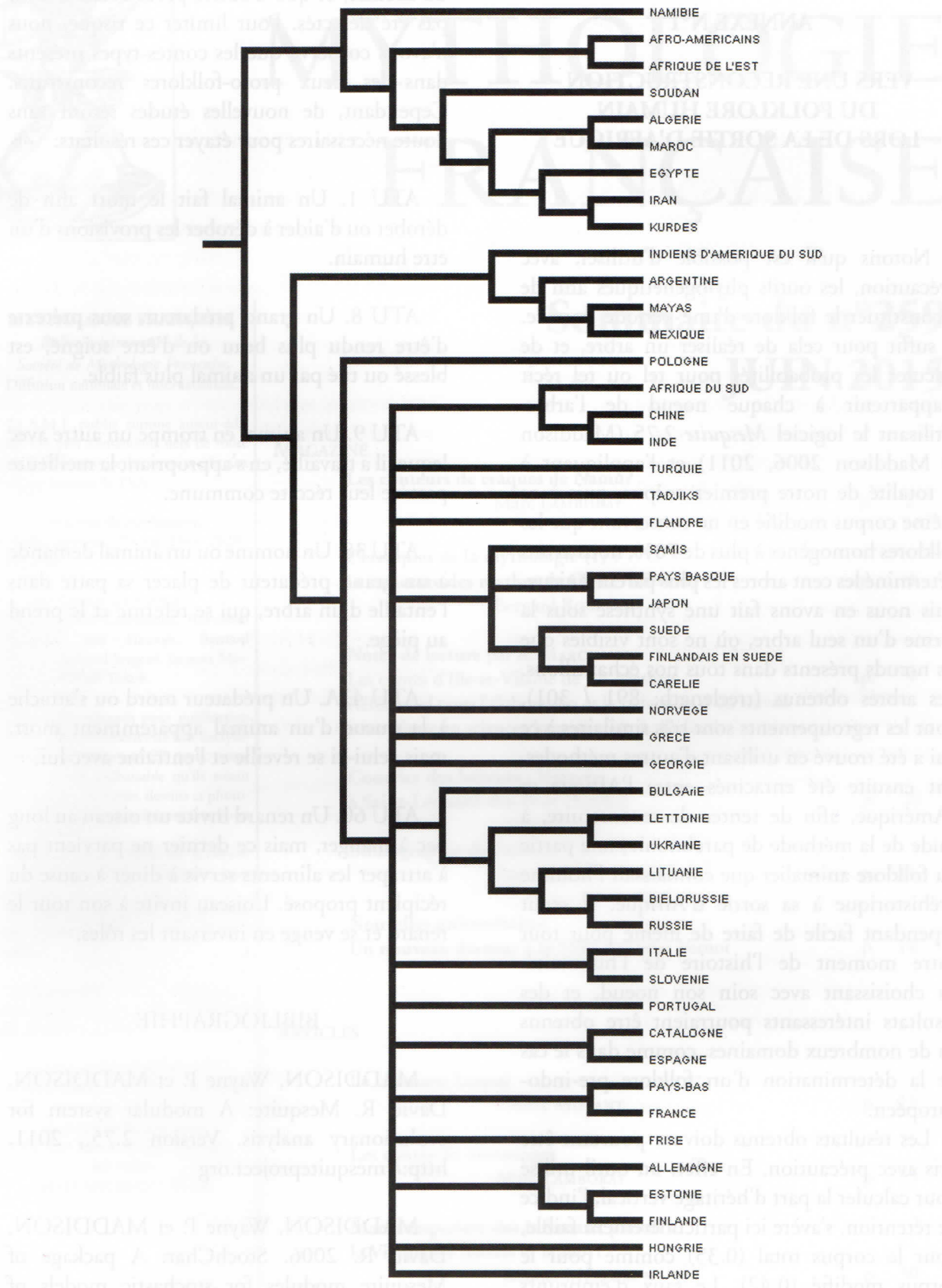
THOMPSON, Stith. *The Types of the Folktales : A Classification and Bibliography; Antti Aarne's « Verzeichnis der Märchentypen », Translated and Enlarged*. Helsinki : FFC 184, 1961.

TORRONI, Antonio, BANDELT, Hans-Jürgen, MACAULAY, Vincent, et al. A signal, from human mtDNA, of postglacial recolonization in Europe. *The American Journal of Human Genetics*, 2001, vol. 69, no 4, p. 844-852.

UTHER, Hans-Jörg. *The Types of International Folktales. Part I: Animal Tales, Tales of Magic, Religious Tales and realistic Tales, with an Introduction*. Helsinki, Academia Scientiarum Fennica, 2004.

UTHER, Hans-Jörg. The Fox in World Literature: Reflections on a "Fictional Animal". *Asian folklore studies*, 2006, p. 133-160.

WITZEL, Michael. *The origins of the world's mythologies*. Oxford, Oxford University Press, 2012.



Arbre de consensus rassemblant les cent arbres les plus parcimonieux. La méthode SPR (subtree pruning and regrafting) a été utilisée pour tester la disposition de l'arbre dans l'espace ; cet algorithme heuristique permute des parties de l'arbre, en cherchant à obtenir le meilleur arbre possible.

ANNEXE N°1

VERS UNE RECONSTRUCTION
DU FOLKLORE HUMAIN
LORS DE LA SORTIE D'AFRIQUE

Notons qu'il est possible d'utiliser, avec précaution, les outils phylogénétiques afin de reconstituer le folklore d'une période donnée. Il suffit pour cela de réaliser un arbre, et de calculer les probabilités pour tel ou tel récit d'appartenir à chaque noeud de l'arbre. Utilisant le logiciel *Mesquite* 2.75 (Maddison et Maddison 2006, 2011) et l'appliquant à la totalité de notre premier corpus, puis à ce même corpus modifié en ne conservant que les folklores homogènes à plus de 70%, nous avons déterminé les cent arbres les plus parcimonieux, puis nous en avons fait une synthèse sous la forme d'un seul arbre, où ne sont visibles que les nœuds présents dans tous nos échantillons. Les arbres obtenus (treelength: 891 / 301), dont les regroupements sont très similaires à ce qui a été trouvé en utilisant d'autres méthodes, ont ensuite été enracinés entre l'Afrique et l'Amérique, afin de tenter de reconstruire, à l'aide de la méthode de parcimonie, une partie du folklore animalier que connaissait l'homme préhistorique à sa sortie d'Afrique. Il serait cependant facile de faire de même pour tout autre moment de l'histoire de l'humanité, en choisissant avec soin son noeud, et des résultats intéressants pourraient être obtenus en de nombreux domaines, comme dans le cas de la détermination d'un folklore pre-indo-européen.

Les résultats obtenus doivent pourtant être pris avec précaution. En effet, un outil utilisé pour calculer la part d'héritage vertical, l'indice de rétention, s'avère ici particulièrement faible, pour le corpus total (0.35) comme pour le corpus modifié (0.42). Le taux d'emprunts étant très élevé, comme nous l'avons vu dans l'article, on ne peut exclure que certains récits reconstruits ici ne soient qu'un artefact dû à l'influence du folklore occidental sur le reste

du monde, et que d'autres proto-récits n'aient pas été détectés. Pour limiter ce risque, nous n'avons conservé que les contes-types présents dans les deux proto-folklores reconstruits. Cependant, de nouvelles études seront sans doute nécessaires pour étayer ces résultats:

ATU 1. Un animal fait le mort afin de dérober ou d'aider à dérober les provisions d'un être humain.

ATU 8. Un grand prédateur, sous prétexte d'être rendu plus beau ou d'être soigné, est blessé ou tué par un animal plus faible.

ATU 9. Un animal en trompe un autre avec lequel il a travaillé, en s'appropriant la meilleure part de leur récolte commune.

ATU 38. Un homme ou un animal demande à un grand prédateur de placer sa patte dans l'entaille d'un arbre, qui se referme et le prend au piège.

ATU 47A. Un prédateur mord ou s'attache à la queue d'un animal apparemment mort, mais celui-ci se réveille et l'entraîne avec lui.

ATU 60. Un renard invite un oiseau au long bec à manger, mais ce dernier ne parvient pas à attraper les aliments servis à dîner à cause du récipient proposé. L'oiseau invite à son tour le renard et se venge en inversant les rôles.

BIBLIOGRAPHIE

MADDISON, Wayne P. et MADDISON, David R. *Mesquite: A modular system for evolutionary analysis*. Version 2.75., 2011, <http://mesquiteproject.org>

MADDISON, Wayne P. et MADDISON, David R. 2006. *StochChar: A package of Mesquite modules for stochastic models of character evolution*. Version 1.1., 2006.