

1 **Inferring patterns of folktale diffusion using genomic data**

2 **SI Appendix**

3 Eugenio Bortolini, Luca Pagani, Enrico R. Crema, Stefania Sarno, Chiara Barbieri, Alessio Boattini, Marco
4 Sazzini, Sara Graça da Silva, Gessica Martini, Mait Metspalu, Davide Pettener, Donata Luiselli, Jamshid J.
5 Tehrani

1	Contents	
2	<u>1 EXTENDED DATASET DESCRIPTION</u>	3
3	<u>2 DISTANCES</u>	3
4	2.1 SNP FILTERING	3
5	2.2 GENETIC, FOLKTALE, ETHNOLINGUISTIC AND GEOGRAPHIC DISTANCES	4
6	2.2.1 GENETIC DISTANCE	4
7	2.2.2 FOLKTALE DISTANCE	4
8	2.2.3 GEOGRAPHIC DISTANCE	5
9	2.2.4 EUCLIDEAN DISTANCES	5
10	<u>3 SPACEMIX.....</u>	5
11	<u>4. NEIGHBORNET.....</u>	12
12	<u>5. AMOVA</u>	13
13	<u>6. BIAS-CORRECTED DISTANCE CORRELATION AND PARTIAL DISTANCE CORRELATION</u>	14
14	<u>7. EXPLORING ASSOCIATION BETWEEN VARIABLES OVER CUMULATIVE GEOGRAPHIC DISTANCE</u>	15
15	<u>8. DIFFUSION OF MOST POPULAR TALES AND ESTIMATION OF POSSIBLE FOCAL POINTS FROM SPATIAL</u>	
16	<u>DISTRIBUTION</u>	16
17		
18		

1 **1 Extended dataset description**

2 Folktale data were sourced from the Aarne Thompson Uther (ATU) Index - a catalogue of over 2,000
3 distinct “international tale types” distributed among more than 200 cultures [22]. Each international
4 type represents an independent, self-contained storyline comprising a combination of motifs (e.g.
5 specific events, characters, or artefacts) that is recognizably stable across cultures. We constructed a
6 dataset recording the cross-cultural distributions of two groups of folktales: 'Animal Tales' (ATU 1 – 299),
7 which feature non-human protagonists, as typified by Aesop's fables, and 'Tales of Magic' (ATU 300 –
8 749), which concern beings or objects with supernatural powers, such as fairies, witches or magic rings
9 [22]. We focused on these two genres because they are the most richly documented and most culturally
10 widespread groups of tales in the ATU Index.

11 73 of the 198 societies in which the tales were recorded could be matched with populations for which
12 whole genome sequences were available (Table S2-I). Of these, 33 (Dataset_{MAIN}) were selected based on
13 a threshold of minimum richness (i.e. those exhibiting at least 5 folktales; **Table S2-II**) and the presence
14 of viable genetic proxies. Each population was univocally described by a string listing the presence (1) or
15 absence (0) of any of the included 596 folktales (**Table S2-II**).

16 In addition to Dataset_{MAIN} we generate an additional subset which is functional to testing explicit
17 hypotheses, i.e. Dataset_{EURASIA} (N=30) which does not include the 3 African population present in
18 Dataset_{MAIN} (Table S1-II, i.e. Congolese, Tanzanian, and West African);

19 **2 Distances**

20 **2.1 SNP filtering**

21 The whole genome sequences used in this study were generated, QCed and phased as part of a broader
22 study [21]. The bulk of ~39M SNPs were used to calculate the statistics described below.

23

1 2.2 Genetic, Folktale, Ethnolinguistic and Geographic distances

2 2.2.1 Genetic distance

3 Genetic distances were estimated by the average pairwise distances between two genomes, one from
4 each population. Genetic distance for (i,j) pairs of populations represented by more than one genome
5 each was calculated as the average of all possible (i,j) pairs of genomes. As a consequence the diagonal
6 of the genetic distance matrix was not constrained to be zero (Table S2-3.1-3).

7 2.2.2 Folktale distance

8 Since the original dataset (Table S1-I) and Dataset_{MAIN} (Table S1-II) comprise binary evidence of presence
9 (1) or absence (0) of a given folktale in a set of worldwide populations, we calculate folktale distance
10 between populations as an asymmetric pairwise Jaccard distance². Symmetric Jaccard distance between
11 population A and population B is calculated as

$$12 \quad J_{\delta}(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|} \quad (1)$$

14 in other words as the ratio between the number of differences and the sum of similarities and
15 differences which can be identified by comparing A and B. In the present work, we assume $X_{ij} = 0$
16 (absence of the j th tale in the i th population of dataset X) to be the ancestral state. Accordingly, we
17 adopt an asymmetrical coefficient that does not consider absence of the j th tale in two sampled
18 populations i and k ($X_{ij} + X_{kj} = 0$) as an instance of homology (for the substantial limits posed by double
19 zeros to inference in ecology and related disciplines please refer to Legendre and Legendre³).

20 Therefore, in a dataset X formed by n rows each representing a population univocally described by a
21 string of presence (1)/absence (0) values of J folktales, we eliminate double zeros and calculate pairwise
22 folktale distance (F_{δ}) between population X_i and population X_k as

$$23 \quad F_{\delta}(X_i, X_k) = \frac{\sum_{\substack{j=1 \\ k \neq i}}^J [X_{ij} + X_{kj} = 1]}{\sum_{\substack{j=1 \\ k \neq i}}^J [X_{ij} + X_{kj} = 1] + \sum_{\substack{j=1 \\ k \neq i}}^J [X_{ij} + X_{kj} = 2]} \quad (2)$$

24 where square Iverson brackets equal 1 if their internal condition is satisfied and 0 if it is not satisfied. The
25 resulting value is the ratio between the number of inter-population differences, and the sum of inter-
26 population differences and similarities based on the presence of the j th folktale in both populations.

1 **2.2.3 Geographic distance**

2 Geographic distances were calculated as pairwise Great Circle Distance using the package *gdistance* in R
3 [42] and by constraining the hypothesised movement of people through one waypoint located in the
4 Sinai Peninsula. Coordinates (longitude and latitude in decimal degrees) expressing the location of each
5 population comprised in Table S2-5 identify the assumed centre of the area occupied by a given folkloric
6 tradition as defined by ATU index.

7 **2.2.4 Euclidean distances**

8 In order to perform bias corrected and partial distance Correlation, folktale, genetic, and geographic
9 distances were transformed into their exact Euclidean representations (as indicated in Szekely et al
10 2007, 2013). The original folktale and genetic distance matrices were scaled through Classic
11 Multidimensional Scaling using the function *cmdscale* in R and following the procedure for exact
12 representation presented by Szekely et al. (2013a). Euclidean distances were computed from the
13 obtained n-2 number of descriptors using the function *dist* in R with method set to “euclidean”.
14 Euclidean representation of geographic distance was instead obtained by reprojecting the original set of
15 coordinates on a plane using two-point equidistant projection through the function *spTransform* in the
16 package *sp* in R (Pebesma and Bivand 2005; Bivand et al. 2013) . Actual Euclidean distance between the
17 new set of coordinates was computed using the function *radish* in the package *fields* in R (Nychka et al.
18 2016).

19

20 **3 SpaceMix**

21 We performed two independent SpaceMix (Bradburd et al. 2013) analyses aimed at retrieving the
22 “genetic” and the “folktale” spatial positions of our Eurasian samples. For the genetic run, we used the
23 126,554 SNPs of chromosome 22 that were variable in at least one of 30 samples each representing one
24 of our studied populations. Given the high number of available markers we deemed a single
25 chromosome to be sufficient to yield reliable results.

26 The geographic information was obtained from Table S2-5.1, the genetic information was inputted using
27 the count/total option and Spacemix was used with the default parameters:

28 `run.spacemix.analysis(n.fast.reps=10, fast.MCMC.ngen=1e5, fast.model.option="target",`
29 `long.model.option="source_and_target", data.type="counts", sample.frequencies = NULL,`

1 mean.sample.sizes = NULL, counts = count, sample.sizes = total, sample.covariance = NULL,
2 target.spatial.prior.scale = NULL, source.spatial.prior.scale = NULL, spatial.prior.X.coordinates=coord[,1],
3 spatial.prior.Y.coordinates=coord[,2], round.earth=FALSE, long.run.initial.parameters = NULL,
4 k=nrow(count), loci=ncol(count), ngen=1e6, printfreq=1e2, samplefreq=1e3, mixing.diagn.freq = 50,
5 savefreq=1e5, directory = NULL, prefix = "OurPrefix")

6 This yielded a “geno-geographic” positioning for each of the 30 samples.

7 The same procedure was replicated using this time “folktale” information as input data. Particularly we
8 generated a pseudo-genetic file where each population was typed as a single individual and the
9 presence of a given folktale was registered as an homozygous trait.

10 The “geno-geographic” and “folk-geographic” coordinates hence generated were compared with the
11 actual geographic coordinates, denoting a tendency for the folk-geographic coordinates to approximate
12 better than the geno-geographic ones the actual geographic locations of the sampled populations
13 (Figure S1-3.1).

14

15 **Figure S1-3.1** SpaceMix analyses for each of the Eurasian populations showing the geographic (dot),
16 geno-geographic (G) and folk-geographic (F) coordinates joined by white segments.

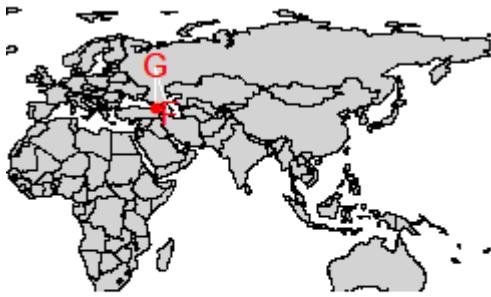
17

18

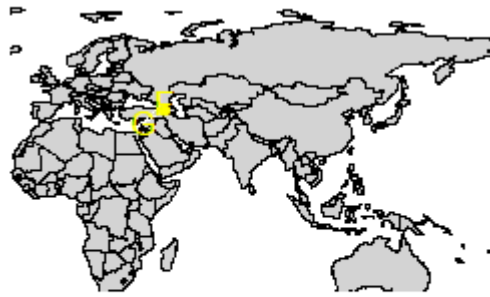
19

20

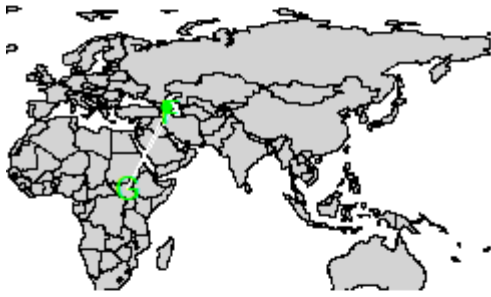
Abkhaz



Armenian



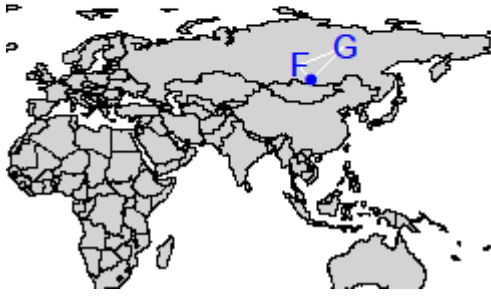
Azerbaijani



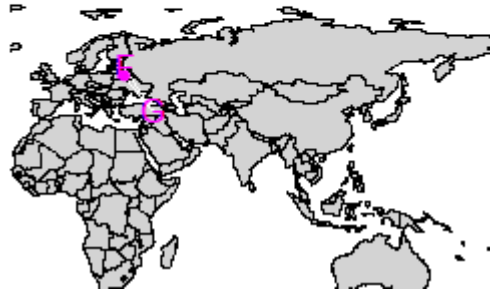
Burmese

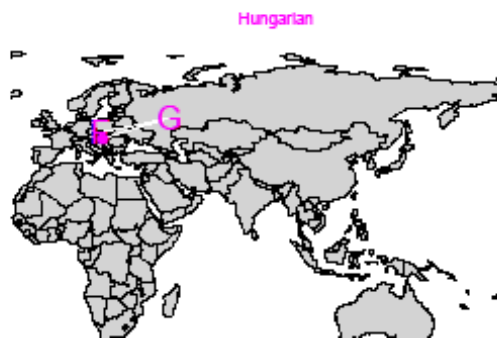
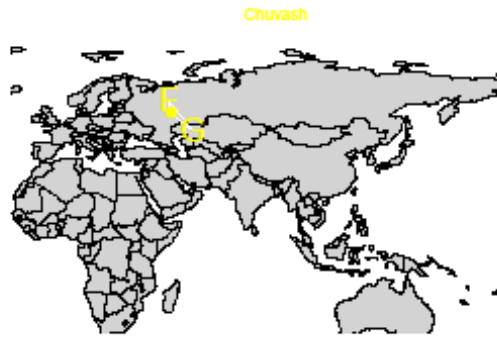


Buryat



Byelorussian





1

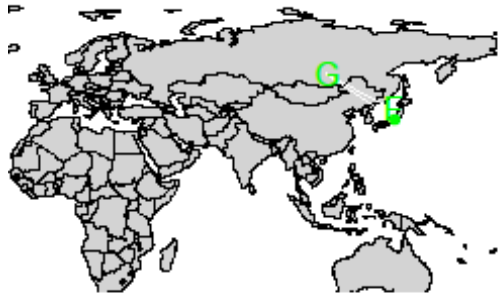
Iranian



Italian



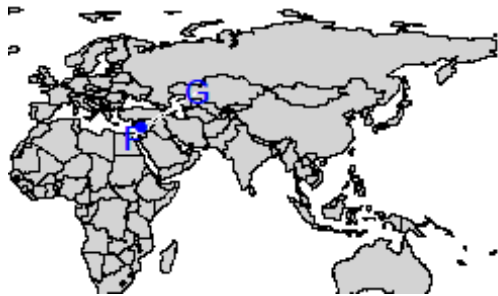
Japanese



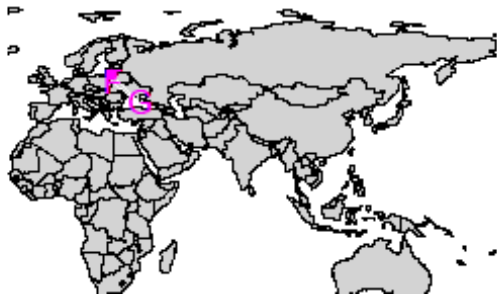
Jordanian



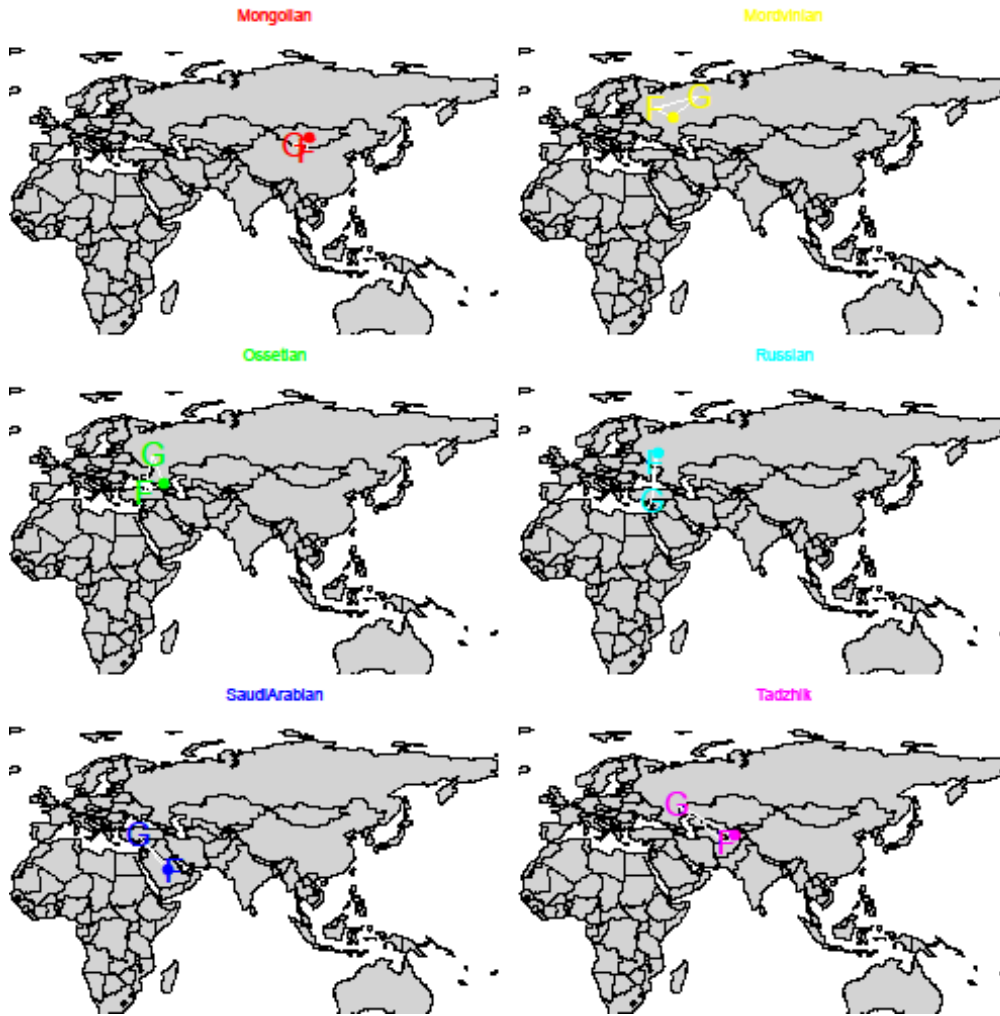
Lebanese



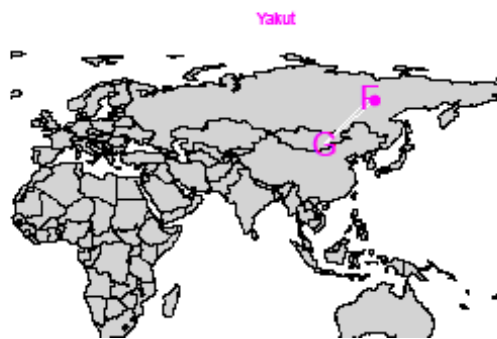
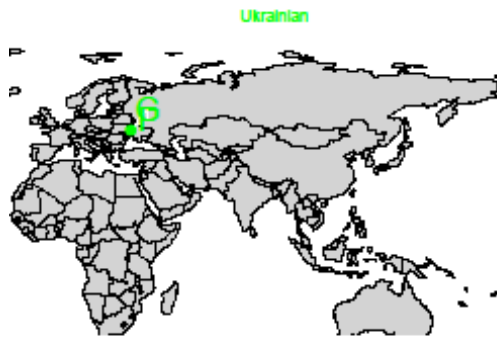
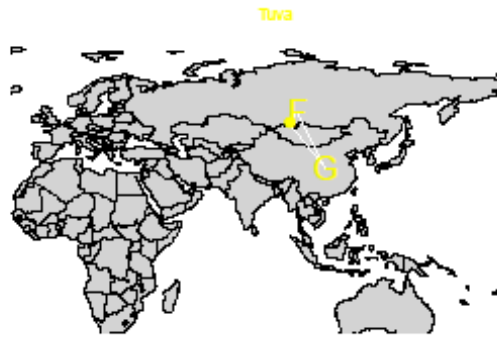
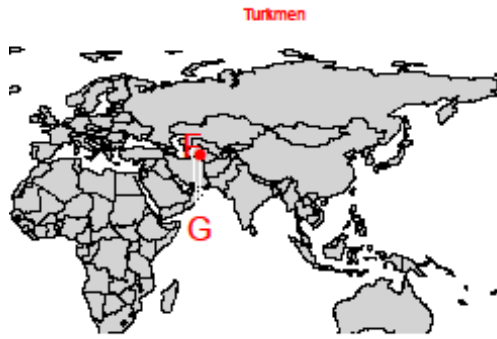
Lithuanian



1



1



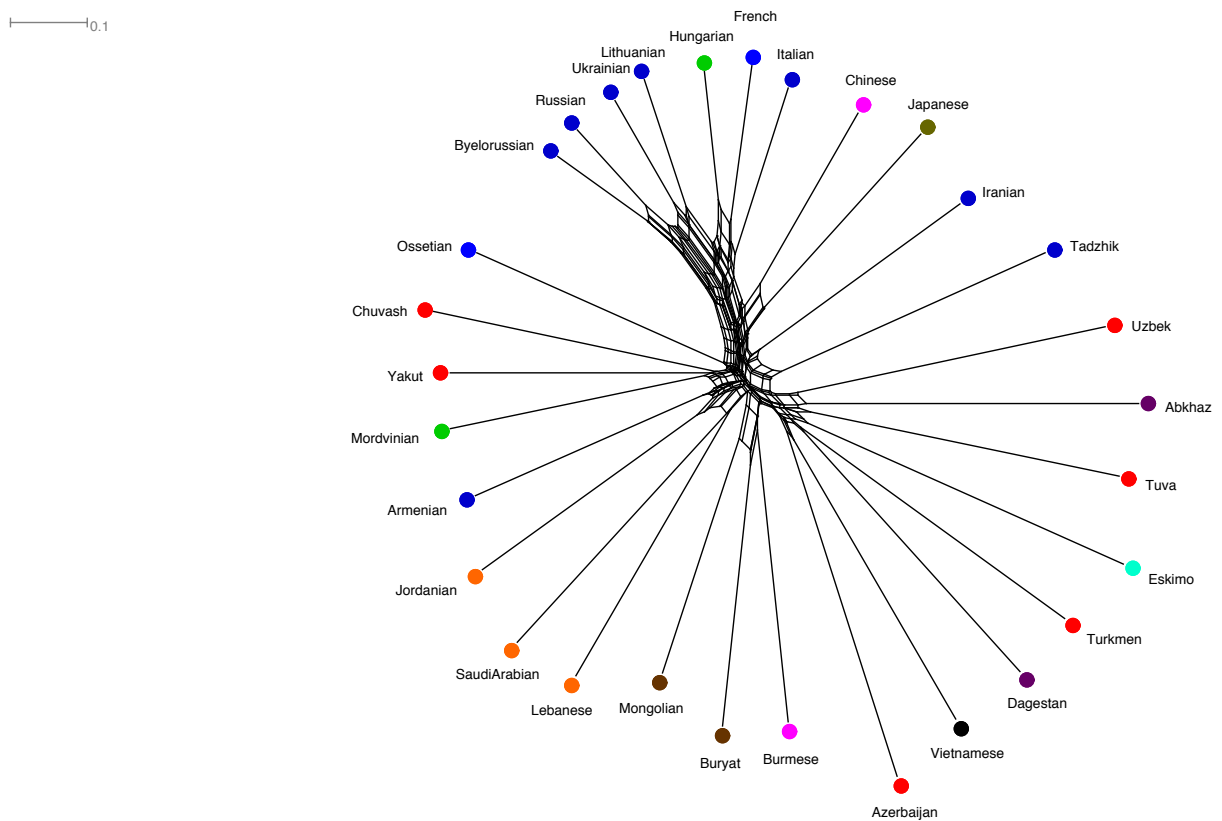
1

2

1 **4. NeighborNet**

2 A NeighborNet analysis was also carried out to further explore the impact of geography and linguistic
3 ancestry on the distribution of folktales in the present dataset. The analysis yielded the following graph
4 (Fig. S1-4.1), exhibiting a certain degree of spatial clustering in addition to proximity and reticulation
5 among linguistic close relatives (e.g. within the Indo-European family, the Semitic family, and the
6 Mongolian family). Overall, the spatial structure in the dataset seems to be stronger (e.g. the position of
7 Hungarian, Japanese/Chinese). Some language families, notably Turkic, Uralic and Caucasian, are
8 scattered across the network. The degree of reticulation is quite high, as testified by the relevant quartet
9 statistics (delta score = 0.317; Q-residual score = 0.002335), suggesting that cultural admixture processes
10 between demes may have an important role.

11 **Fig. S1-4.1. Neighbornet graph based on folktale distance. Linguistic color code: red = Turkic; blue =**
12 **Indo-European; pink = Sino-Tibetan; purple = Caucasian; turquoise = Eskimo-Aleut; orange = Semitic;**
13 **light green = Uralic; dark green = Japonic, brown = Mongolian; black = Austro-asiatic**



14

15

1 5. AMOVA

2 We use Analysis of Molecular Variation (AMOVA; Excoffier et al. 1992) to formally assess the impact of
3 ethnolinguistic boundaries on both genetic and cultural (folktale) variability. To do this, we assigned each
4 population to an ethnolinguistic group (derived from Ethnologue; SI table S2-9.1), and used the function
5 `amova` in the package `pegas` in R to run the analysis (Paradis 2010). AMOVA is commonly used in
6 population genetics to assess the degree of population structure in a metapopulation. In other words, it
7 measures the degree of variability existing between predetermined groups as opposed to the amount of
8 variability observed within each group. AMOVA returns a summary statistic (Φ_{ST}) obtained by
9 computing the ratio between intergroup diversity estimate and total diversity in the metapopulation.
10 Although it is derived from the more general class of F_{ST} measures, Φ_{ST} evaluates symmetric distance
11 matrices while F_{ST} is based on correlations between individual variant frequencies. Such measures have
12 already been successfully adopted to investigate co-evolutionary patterns in genetic and cultural
13 datasets (Bell et al. 2009; Rzeszutek et al. 2012), in cultural datasets alone (Shennan et al 2015), and in
14 one case on the distribution of folktale variants in Europe (Ross et al. 2013). Results of these works
15 consistently show that average intergroup cultural dissimilarity is stronger than genetic dissimilarity
16 measured on the same set of demes, while Φ_{ST} values obtained for cultural markers are usually in a
17 range comprised between 0.02 (musical diversity; Rzeszutek et al. 2012) and higher levels for variants of
18 individual folktales ($\Phi_{ST} = 0.09$; Ross et al. 2013), or personal ornaments ($\Phi_{ST} = 0.109$) and pottery
19 ($\Phi_{ST} = 0.134$) in Neolithic Europe (Shennan et al 2015). Our results confirm this differential impact on
20 genetic variability on the one hand, and cultural variability on the other.

21

22

23

24

25

26

27

28

29

30

31

32

33

34

1 **6. Bias-corrected distance correlation and partial distance correlation**

2
3 Distance correlation is a measure of statistical dependence between two variables which is specifically
4 suited for testing such hypothesis on pairs of symmetric distance matrices. Its value equals zero if and
5 only if the two variables are statistically independent (Székely et al. 2007). In addition: a) the resulting
6 statistics are not bound to linear models. On the contrary, they are sensitive to all types of dependent
7 relationships, including nonlinear and non monotone ones (Székely et al 2007); b) it is not prone to the
8 same problems raised for standard and partial Mantel tests (Guillot and Rousset 2013); and c) it is
9 usually preferred to other measures of nonlinear association when small or practical sample size are
10 concerned (Gorfin et al 2011).

11 Given two distance or dissimilarity matrices, standard distance correlation performs double centering of
12 each matrix by subtracting row and column average to rows and columns of the original matrix, and
13 adding the grand mean of the distance matrix to the results, so that all columns and rows of the resulting
14 matrices sum to zero. Distance correlation between two such scaled matrices - as in Pearson's Product-
15 moment correlation coefficient is computed by dividing the distance Covariance by the product of the
16 respective distance standard deviations. Distance Covariance is obtained by computing the summed
17 cross-product between the two double-centered matrices and averaging it over squared sample size.
18 Distance correlation is therefore not the correlation between original distances. It is instead based on
19 cross-products between scaled moment obtained by double-centering the original matrices.

20 In the present paper, we perform Bias-Corrected Distance Correlation (Szekely and Rizzo 2013) suited for
21 bigger sample size (in the present study we have 435 observation when all pairs are considered over 30
22 populations, which is exactly the example size provided by the authors and developers of the method),
23 This method corrects for potential limitations of original distance Covariance and distance Correlation
24 measures (Székely et al. 2007) when dimension tends to infinity, and is based on an unbiased estimator
25 or the squared distance population covariance. A suited t-test of independence is offered. In addition,
26 we perform Partial distance correlation (Szekely et al 2013a) to assess the impact of one variable over
27 another, while controlling for the effect of a third variable. For calculating partial distance correlation the
28 standard double-centering used in standard and bias-corrected distance correlation is replaced by a
29 different centering technique named U-centering and based on the demonstration that such
30 transformed distance and dissimilarity matrices have a corresponding U-centered Euclidean
31 representation in Hilbert space (a generalization of Euclidean plane with a finite or infinite number of

1 dimensions; Szekely et al 2013a). Simulation studies confirm that the joint significance test controls type
 2 I error rate at its nominal level, and outperforms partial correlation and partial Mantel test in terms of
 3 power (Szekely et al 2013a).

4

5 **7. Exploring association between variables over cumulative geographic** 6 **distance**

7 **Table S1-7.1 Model comparison over cumulative geographic distance. Results report Pearson's**
 8 **product-moment correlation coefficients plotted in Fig.2 and obtained with original distance matrices.**

N	bindist	r.folk geo	p.folk geo	r.folk(Lw) geo	p.folk(Lw) geo	r.folk geno	p.folk geno	r.folk(Lw) geno(Lw)	p.folk(Lw) geno(Lw)
115	2000	0.16	0.09	0.21	0.03	0.45	<0.001	0.40	<0.001
249	4000	0.24	<0.001	0.58	<0.001	0.34	<0.001	0.40	<0.001
343	6000	0.22	<0.001	0.68	<0.001	0.23	<0.001	0.51	<0.001
412	8000	0.22	<0.001	0.67	<0.001	0.22	<0.001	0.55	<0.001
434	10000	0.19	<0.001	0.64	<0.001	0.20	<0.001	0.55	<0.001
435	12000	0.19	<0.001	0.64	<0.001	0.20	<0.001	0.55	<0.001

9

10

11

12

13

14

15

16

17

1 **8. Diffusion of most popular tales and estimation of possible focal points** 2 **from spatial distribution**

3
4 We identify 19 “most popular” tales, i.e. folktales that are present in at least 30 populations out of 60
5 Old World populations available in the original presence/absence matrix (Table S2-10, S2-11), which are
6 briefly summarized below:

- 7
- 8 ○ **ATU 155 ‘The Ungrateful Snake Returned to Captivity’**: A snake (or another dangerous animal)
9 attacks a man who rescues it, and is punished by other animals.
 - 10 ○ **ATU 300 ‘The Dragon Slayer’**: A man rescues a beautiful maiden from a dragon/monster, often
11 with the help of his dogs. Later, he exposes an imposter who claims credit for the deed.
 - 12 ○ **ATU 301 ‘The Three Stolen Princesses’**: A man rescues three women from a pit. His companions
13 betray him by leaving him in the pit and stealing the girls. With the aid of a spirit the hero flies
14 up and exposes his companions, marrying the youngest girl.
 - 15 ○ **ATU 303 ‘The Twins, Or Blood Brothers’**: A hero rescues a woman and marries her, but is later
16 bewitched. His twin brother sets out to find him, and is mistaken by the woman for her husband.
17 The twin releases his brother, who kills him in a jealous rage, mistakenly believing him to have
18 seduced his wife. The twin is later resuscitated.
 - 19 ○ **ATU 313 ‘The Magic Flight’**: A man elopes with the daughter of a demon or king. She uses
20 magical objects to obstruct their pursuers and they escape.
 - 21 ○ **ATU 314 ‘Goldener’**: A golden-haired man marries the king’s daughter. He is mocked by his
22 brothers-in-law, but succeeds in completing heroic deeds where they fail and is made the heir.
 - 23 ○ **ATU 325 ‘The Magician and his Apprentice’**: A boy is given to a magician to be his apprentice.
24 The boy learns the art of sorcery and frees himself from his master after a battle in which they
25 transform into a succession of different kinds of animals.
 - 26 ○ **ATU 400 ‘The Man on a Quest for his Lost Wife’**: A miscellaneous group of stories concerning a
27 man who is separated from his wife during an adventure. When he finds her she is about to
28 marry another man, but he proves his identity to her and they are reconciled.

- 1 ○ **ATU 403 'The Black and the White Bride'**: A girl is to marry the king, but her stepmother tries to
2 kill her and replace her with her own daughter. The girl proves her identity to the king and
3 exposes her stepsister as an imposter.
- 4 ○ **ATU 480 'The Kind and Unkind Girls'**: A girl goes on a journey and is kind to those she
5 encounters. She is rewarded. Her stepmother sends her own daughter on the same journey but
6 she is unkind and gets punished.
- 7 ○ **ATU 531 'The Clever Horse'**: A miscellaneous group of stories in which a young man is helped to
8 complete some near-impossible tasks by a talking horse and marries a princess.
- 9 ○ **ATU 550 'Bird, Horse and Princess'**: Three brothers are sent on a quest by their father to catch a
10 magic bird. The youngest brother succeeds but is betrayed by the other two who try to claim the
11 prize. With the help of a magical animal the hero exposes his brothers.
- 12 ○ **ATU 554 'The Grateful Animals'**: A man helps a series of animals, who reciprocate by helping
13 him to complete a series of near-impossible tasks.
- 14 ○ **ATU 560 'The Magic Ring'**: A boy acquires a magic ring that grants him wishes. He marries a
15 princess, who steals the ring to elope with her lover. The boy recovers a ring and punishes his
16 faithless wife and her lover.
- 17 ○ **ATU 563 'The Table, the Donkey and the Stick'**: A man acquires magical objects from a
18 supernatural being. He is cheated out of them and given plain objects in their place, but
19 manages to recover his possessions and punish the cheat.
- 20 ○ **ATU 613 'The Two Travellers'**: After losing an argument with his companion, a man is blinded.
21 He learns the secrets of birds and recovers his sight as well as gaining new powers. His
22 companion imitates him and is punished by the birds.
- 23 ○ **ATU 670 'The Man Who Understands Animal Languages'**: A snake teaches a man the languages
24 of animals on condition he keeps it a secret. The man's wife nags him to teach her but he refuses
25 after being warned of the consequences by a male animal (usually a rooster).
- 26 ○ **ATU 700 'Thumbling'**: A couple wish for a child and are given a tiny boy through supernatural
27 means. The boy is lost and goes on a series of adventures until he is reunited with his parents.
- 28 ○ **ATU 707 'The Three Golden Children'**: A woman marries a king and gives birth to three children,
29 who are stolen by her jealous sisters. When they grow up the children go on a quest to find their
30 parents and eventually expose the sisters.

1 One possible explanation for the wide dispersion of these tales is that they spread through the
2 dissemination of written texts, which would allow them to travel much further and in a much shorter
3 period than would be possible solely through the vectors of human dispersal and traditional oral
4 transmission. Such a process would most likely have coincided with the emergence of the fairy tale as a
5 popular literary genre in the sixteenth and seventeenth centuries (Bottigheimer 2014). Although many
6 folktales were incorporated into literary works prior to this period (for example, in medieval romances),
7 the development of new, cheap printing technologies together with the growth of international trade
8 networks and European colonialism would have allowed tales to circulate to much wider audiences than
9 was previously possible (ibid.).

10 In fact, seven of the tales listed above were published in major fairy tale collections during this period,
11 including Giovanni Francesco Straparola's *Le Piacevoli Notti* in 1550-55 (ATU 314, ATU 325, ATU 670),
12 Giambattista Basile's *Lo cunto de li cunti* in 1634 (ATU 301, ATU 480, ATU 560), Charles Perrault's
13 *Histoires ou contes du temps passé* in 1697 (ATU 480, ATU 700). However, one obvious question raised
14 by the hypothesis that these tales spread via textual transmission is why the other stories contained
15 within these collections did not achieve similarly wide cross-cultural distributions. Secondly, while
16 literary versions have undoubtedly made a major contribution to the modern forms of these tales, there
17 is compelling evidence that they were derived from already well-established and widespread oral
18 traditions, rather than the other way round (Ben-Amos et al. 2010). For example, ATU 301 'The Three
19 Stolen Princesses' occurs in Greek and Indian myths that long predate Basile's Italian fairy tale of 1634.
20 Similarly, versions of ATU 325 'The Magician and His Pupil', ATU 560 'The Magic Ring' and ATU 670 'The
21 Man Who Understands Animal Languages' appear in Indian and Middle Eastern sources (including the
22 Ramayana and One Thousand and One Nights) that are clearly independent of later European literary
23 versions of these tales (Thompson 1977).

24 An alternative explanation for the distribution of these tales is that they reflect signatures of demic or
25 cultural diffusion that are more ancient than the other patterns detected in the dataset. In order to
26 characterize these signatures, we sought to identify possible centers of origin and dispersal for the tales.
27 To do so, we assessed the amount of linear correlation between geographic distance from each
28 population exhibiting a given tale and the distribution of the proportion of the remaining populations
29 displaying that tale over the same geographic gradient. More specifically, we binned pairs of populations
30 into fixed intervals of geographic distance (2000 Km), and for each bin we calculated the proportion of
31 populations exhibiting a given folktale. We then calculate linear correlation between the distribution of

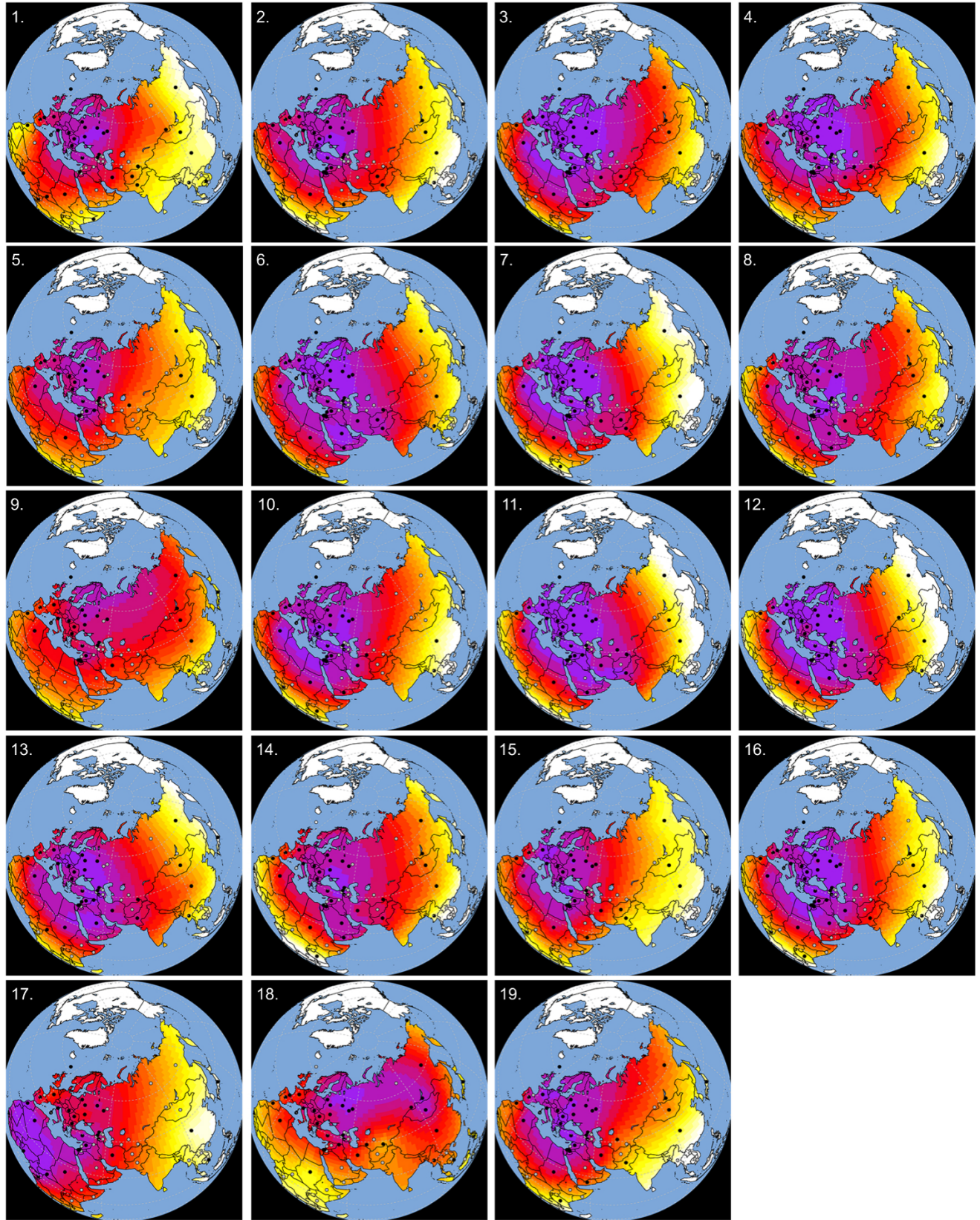
1 percentages obtained for each geographic bin and geographic distance from all the populations in the
2 dataset that exhibit that folktale. All the above mentioned analyses have been performed in R. The
3 assumption is that - if we envisage a long-range and ancient diffusion process whose vectors are solely
4 human dispersal and traditional cultural transmission - we expect to obtain a distance-decay patterning
5 with higher percentages indicating potential "origin populations" from which increasingly lower
6 percentages depart forming a clinal trend over geographic distance. To avoid losing information we did
7 not focus on single coordinate pairs exhibiting the highest values for each tale, and plotted instead the
8 distribution of correlation coefficients on a map in order to visually define the most probable area of
9 origin (centres of origin exhibiting the lowest correlation coefficients; Figure S8-I). Probability surfaces
10 were obtained by interpolating correlation coefficients computed for each population using the function
11 producing plate spline interpolation of the *fields* package in R [47].

12 The resulting potential patterns of diffusion are represented in Figure S6-I. Although each of these
13 patterns presents some specificities, some general trends can be found - as summarized in the main text.
14 In particular, four main multi-directional waves of diffusion can be hypothesized:

- 15 1. Potential African origin (e.g. ATU 670)
- 16 2. Southward spread from northern Eurasia (e.g. ATU 700)
- 17 3. Eastern European origin (e.g. ATU 301, 303, 313)
- 18 4. Middle-Eastern/Caucasian origin (e.g. ATU 314, 400, 480, 560)

19
20 While further research is needed to verify these patterns (for example, by reconstructing the
21 evolutionary histories of variants of each tale type to test whether they match the dispersal scenarios),
22 the results have significant implications for current understandings about the origins of international
23 folktale traditions. In particular, they suggest a less Euro-centric view of tale origins than traditional
24 "historic-geographic" reconstructions based on the frequency of variants (i.e. the number of versions of
25 a given tale type recorded in each population) and chronology of literary versions. A major problem with
26 this approach is that conclusions about a tale's origins may often be skewed by the strong European bias
27 in both the richness of the folktale and literary records. For example, ATU 300 'The Dragon Slayer' –
28 which has been proposed to be the original archetype storyline from which all fairy tales are derived
29 (Propp 1968)– was previously believed to have originated in medieval western Europe, most likely
30 France, where the earliest known versions were recorded. Our analysis instead suggests that this tale –
31 together with the related tale ATU 313 'The Twins' may have arrived in western Europe from further

1 east, either from the region of modern day Ukraine and Belarus, or of Turkey and Kurdistan. Similarly,
2 whereas folklorists have claimed that ATU 670 'The Man Who Understands Animal Languages' originated
3 in Europe and was transported to Africa through colonialism, our findings reverse the direction of
4 transmission and suggest that the tale probably arose in North Africa.



1 **Figure S1-8-I - Plot of the probable areas of origin of the 19 “most popular” tales:** Probability surfaces
2 have been obtained interpolating correlation coefficients computed for each population. Grey dots
3 indicate populations that do not exhibit the specific tale of interest. 1) Tale 155; 2) Tale 300; 3) Tale 301;
4 4) Tale 303; 5) Tale 313; 6) Tale 314; 7) Tale 325; 8) Tale 400; 9) Tale 403; 10) Tale 480; 11) Tale 531; 12)
5 Tale 550; 13) Tale 554; 14) Tale 560; 15) Tale 563; 16) Tale 613; 17) Tale 670; 18) Tale 700; 19) Tale 707.