

# Supporting Information

## Maurits and Griffiths 10.1073/pnas.1319042111

### Word Order Distribution in Sample

Table S1 shows the number of languages with each basic word order in each of the six language families in our sample.

### Features Used for “Feature” Method

Our “feature” method computes pairwise distances between languages using a total of 10 features from *The World Atlas of Language Structures* (WALS) (1). These features are Order of Adjective and Noun, Preverbal Negative Morphemes, Postverbal Negative Morphemes, Order of Demonstrative and Noun, Negative Morphemes, Position of Interrogative Phrases in Content Questions, Postnominal relative clauses, Definite Articles, Hand and Arm, and Position of negative words relative to beginning and end of clause and with respect to adjacency to verb.

### Model Selection for Different Tree Methods

In *Results*, we presented results based on using the combination method for generating trees and assuming that a single common  $Q$  matrix is shared by all seven families. These decisions are justified on the basis of a Bayesian information criterion (BIC) (2) model selection process. Table S2 shows the maximum log-likelihood and BIC values for all 12 models considered. The BIC takes into account both the goodness-of-fit of a model (as determined by maximized log-likelihood) and the complexity of the model (as determined by the number of free parameters). Lower BIC values indicate better tradeoff between accuracy and complexity.

The common- $Q$  combination approach yields a BIC of 767.06, which is lower than any alternative, which is why we presented that analysis in *Results*. The combination approach also yields the lowest BIC under the individual- $Q$  approach, making its superiority clear. Similarly, for all four tree methods, the common- $Q$  approach yields a lower BIC than the corresponding individual- $Q$  approach, making the superiority of the common- $Q$  approach clear. Note that the individual- $Q$  models do yield higher log-likelihoods, meaning they fit the data better. However, the BIC measure penalizes the drastic increase in free parameters—from 36 to 216. The increased fit to data comes at the cost of excessively increased model complexity. Note that for both common and individual  $Q$  matrices, the feature trees produce extremely poor fits to the data (and correspondingly high BIC values) compared with all other methods.

### Inference Validation

To validate our inference process, we generated many random  $Q$  matrices and used our generative model to create artificial word-order data for the leaf nodes of many of our trees. We then used the inference procedure described in the paper to attempt to recover the parameters. Ancestral word orders were selected at random, while  $Q$  matrices were sampled from the prior distribution, but then had the stability parameter of the root word order set to one of a fixed set of values. This allowed us to examine how our method’s reliability at correctly inferring ancestral word orders depends upon the stability of the true ancestral order.

Fig. S1 shows the proportion of trials for which our inference procedure correctly assigned the highest posterior probability (using a uniform prior) to the word order that was used to generate the leaf data, for various values of the root word-order stability parameter. Sixty trials were performed for each value of the stability parameter, using many trees from different families. When the stability parameter is between 0 and 0.5 (corresponding to a very stable root word order), more than 90% of trials suc-

cessfully recovered the true word order, as expected. For stability parameters higher than 5.0 (corresponding to highly unstable root word orders), performance of the method is around chance (16.66%), once again as expected. The method is more likely to be correct than incorrect so long as the root stability is below around 2.5. For the common- $Q$  inference scheme and combination tree method used in the main text, the mean inferred stabilities of subject (S)–object (O)–verb (V) (SOV), SVO, and VSO were 0.16, 0.19, and 1.34, respectively. For these stability values, our method’s performance is higher than 80%.

### Shuffled Word-Order and Split-Half Reliability

In this section, we perform two tests to assess the reliability of our inference procedure.

**Shuffled Word-Order Testing.** A possible point of concern with our results is that the current number of present-day languages with each word order is contributing more to the results than the phylogenetic trees. To investigate this possibility, we performed an alternate version of our analysis in which the assignments of word orders to leaf nodes of the trees were randomly shuffled. This way, the relative frequency of each word order in the various families does not change, but any relationship between historical relatedness and word order is obscured.

We performed the common- $Q$  combination analysis using shuffled word-order data. This made no difference to the relative stability of the six word orders: When the word orders are ranked by their mean stability parameter values, the same ranking obtains in both cases. However, the absolute stability of all word orders was decreased by shuffling, as we might expect. Shuffling removes the tree structure from the leaf node data, such that neighboring nodes are less likely to have a common word order. Explaining this greater variability requires more-frequent changes of word order.

The maximum likelihood encountered during Markov Chain Monte Carlo (MCMC) sampling is substantially reduced by shuffling, since the model is best able to explain data where closely related languages tend to have similar word order.

The posterior probability of ancestral word orders can also vary substantially after the shuffling, although this depends upon the heterogeneity of the language family. For example, our database includes 89 Sino-Tibetan languages: 81 with SOV word order and 8 with SVO word order. The shuffling operation makes relatively little difference to the analysis of this family: SOV is the most probable ancestor before and after shuffling. For more diverse families, shuffling can drastically change the results. Using the prior family  $P_t$  detailed in the main text, with  $t = 1,000$  y, shuffling changes the posterior distribution over ancestral word orders for Niger-Congo (our most diverse family) from a fairly flat distribution, giving roughly 0.1 to SOV, 0.0 to SVO and 0.2 to all other orders, to one with most of its probability concentrated on SVO (0.25) and VSO (0.55).

These findings indicate that the relative abundance of the different word orders in modern languages and our sampled phylogenetic trees are both making substantial contributions to our results.

**Split-Half Testing.** Another concern is that our results are highly dependent on the particular set of 671 languages that we are using, and would not necessarily generalize to a larger sample. This concern is very reasonable, given that our database represents

less than 10% of the world's current languages and that SVO is overrepresented in our sample.

To investigate this possibility, we perform split-half reliability testing. We take the set of languages in each family and randomly divide it into two equally sized subsets, and from these construct two trees, using the combination method. Each of these two trees represents a random subset of the family. We repeat this process 100 times, ending up with 100 pairs of trees for each language family. We analyze each pair using the common- $Q$  approach and compute the correlations between the posterior mean  $Q$  matrices and ancestral word-order distributions for each tree in the pair. The mean correlation between  $Q$  matrices for a tree pair is 0.99 and, between ancestral word-order posteriors, is 0.92. This suggests our results are not particularly dependent upon the particular set of languages we have sampled.

### Alternative Analysis of Indo-European, Using Extinct Languages

Our analysis of Indo-European in *Results*, when using a uniform prior, has posterior probabilities of 0.23 for both ancestral OVS and OSV. Only a prior preference for stable word orders allows us to find SOV as the most probable ancestor. More evidence for SOV ancestry can be acquired by augmenting our dataset with extinct languages with known word order. Ancient languages correspond to leaf nodes much closer to the root of the language family tree than modern languages, and, as such, information about their word orders is more informative about the ancestral word order.

Pagel's computational analysis of ancestral Indo-European word order (3) makes use of reconstructed SOV word orders for Tocharian and Hittite, two ancient Indo-European languages. Tocharian and Hittite do not exist on the Indo-European trees constructed with any of our four methods, because the necessary data are not available in WALS (1). However, the reference Indo-European trees we used to calibrate our methods do include leaves for these languages. We perform analysis of this tree, using the posterior mean  $Q$  matrix obtained from our common- $Q$  combination analysis of all six of our families.

First, we compute a posterior over ancestral word orders using only the modern languages in the reference tree. There are 42 such languages, less than the 61 found on our own trees. With a uniform prior, we find posterior probabilities of 0.33 and 0.35 for OVS and OSV: even higher than for our own trees. The posterior probability of ancestral SOV is just 0.05. When we add three SOV leaves to the tree, corresponding to Hittite and two dialects of Tocharian, the results change substantially. The probability of ancestral SOV increases to 0.55, making it the most likely ancestor, while the object-initial word orders drop to 0.18 and 0.19, a little more than half their previous values. This demonstrates the strong impact that knowledge of ancient languages can have on analysis of ancestral word orders. Unfortunately, ancient languages for which we have reliable word-order knowledge are very scarce.

### Priors for Ancestral Word Orders

The uniform prior analysis presented in *Results* is surprising in that it results in high posterior probabilities of VOS, OVS, and OSV being ancestral word orders for many of the language families considered. As explained, this is due to the very low stability of these word orders. This shortcoming can be overcome using a prior distribution, and we presented the results of one family of priors in *Results*. However, other priors are certainly possible and worthy of investigation.

**Stationary Distribution.** One of these alternatives is to use the stationary distribution of the Markov chain defined by  $Q$ . Like the prior family used in *Results*, this approach penalizes unstable word orders. However, it also strongly favors the most stable

word orders over only slightly less stable word orders. The stationary distribution has almost all of its mass on the subject-initial word orders (0.41 for SOV and 0.52 for SVO). This prior is so strong that it can potentially "swamp" the signal from the data.

Fig. S3 shows the posterior distributions for the different language families using this prior. The most probable ancestors are VSO for Afro-Asiatic, VOS for Austronesian, SVO for Niger-Congo, and SOV for all other families. This is largely in keeping with the results produced for most priors in the  $P_t$  family, particular for lower values of  $t$  when the penalties against VSO are not especially strong.

It is difficult to justify the use of the stationary distribution as an ideal prior for this problem. It corresponds to the assumption that the protolanguages for each of our seven families are themselves the very distant descendants of yet earlier languages, which may not be the case.

**Present-Day Distribution.** Another possible prior is the observed present-day distribution of word orders. Like the other priors we have considered, this gives severe penalties to the unstable orders VOS, OVS, and OSV. However, unlike the stationary distribution, it gives only a very slight preference for SOV over SVO. This allows the data to do the majority of the work in choosing between these two possibilities for an ancestral order, rather than being swamped by the prior.

Fig. S4 shows the posterior distributions for this prior. Once again, the most probable ancestors are VSO for Afro-Asiatic, VOS for Austronesian, SVO for Niger-Congo, and SOV for all other families. Compared with the stationary prior, Afro-Asiatic's and Indo-European's ancestral word orders are less ambiguous: Afro-Asiatic has changed from being roughly evenly split between SOV, SVO, and VSO to VSO having a clear lead, and Indo-European has changed from being roughly evenly split between SOV and SVO to SOV having a clear lead.

The present-day distribution is also difficult to justify as an ideal prior. Because the present-day distribution is only slightly different from the inferred stationary distribution, there is no reason to think that it is representative of the distribution at the time of our families' protolanguages.

### Variations of Analysis

The results presented in the main text are based on one version of our analysis, where the phylogenetic trees are generated using the combination method and a single  $Q$  matrix is inferred to fit the data for all six trees simultaneously. A number of variations are possible and have been performed, including using the three other tree-generation methods and permitting each tree to have its own distinct  $Q$  matrix.

**Varying Tree-Generation Methods.** The combination tree-generation method was used for our primary results because it achieved a better fit to the data than any other method. However, a central tenet of our approach is to explore a wide range of trees and see which inferences about language change are robust across that variation. As such, it makes sense to consider even the less accurate tree estimates individually as well. We can place more confidence in any inferences that are robust across all of the sets of trees.

**Word-order dynamics.** Figs. S5, S6, and S7 show the inferred word-order dynamics for the geographic, genetic, and feature tree sets, respectively. These plots, like the one in *Results*, are based on one  $Q$  matrix shared across all families. The geographic and genetic methods produce results broadly compatible with the combination method. The probabilities of SOV and SVO being the most stable word order are 0.59 and 0.41 for geographic, 0.58 and 0.42 for genetic, and 0.69 and 0.32 for combination. SOV is probably more stable than SVO in all cases, but the evidence is never strong (it is strongest with the combination trees). The feature

results differ from all other trees, with mean probabilities of 0.48 and 0.52, i.e., very weak evidence that SVO is more stable than SOV. Recall, though, that the feature trees result in overwhelmingly lower likelihood values for the data than all other methods, so the fact that the feature results are outliers (although only very slightly) in this regard should not disturb our confidence in the other methods.

Regarding preferred directions of change, there is excellent agreement across all four methods: The mean probability that SOV  $\rightarrow$  SVO is preferred over SOV  $\rightarrow$  VSO is 0.93, 0.93, 0.94, and 0.89 for geographic, genetic, feature, and combination methods, i.e., SOV  $\rightarrow$  SVO is clearly preferred in all cases. Similarly, the probabilities that SVO  $\rightarrow$  SOV is preferred over SVO  $\rightarrow$  VSO are 0.94, 0.77, 0.93, and 0.74 for geographic, genetic, feature, and combination methods, i.e., SVO  $\rightarrow$  SOV is preferred in all cases. Finally, the probabilities that VSO  $\rightarrow$  SVO is preferred over VSO  $\rightarrow$  SOV are 0.98, 0.89, 0.99, and 0.94, i.e., VSO  $\rightarrow$  SVO is clearly preferred in all cases. This means that our picture of slow cycles between SOV and SVO with rarer, faster cycles between SVO and VSO is very strongly supported by data. **Ancestral word orders.** Figs. S8, S9, and S10 show the posterior probabilities for different ancestral word orders for the seven families, using the same family of priors as in the main text and based on the geographic, genetic, and feature tree sets, respectively.

Three of the families, Niger-Congo, Sino-Tibetan, and Trans-New Guinea, show qualitatively identical behavior across all four tree sets, with one particular word order (SVO for NC and SOV for ST and TNG) quickly becoming the only order with non-negligible posterior probability. Only the precise values of the prior parameter  $t$  where this happens vary across tree methods.

Recall that the feature trees provide an exceptionally poor fit to the data. If we are willing to disregard these trees and focus only on the remaining three, then Austronesian also displays the same qualitative behavior across different methods. It is most probably VOS for a lower range of  $t$  values and VSO for a higher range, with the crossover point varying with method.

This leaves three families, Afro-Asiatic, Indo-European, and Nilo-Saharan, whose ancestral word-order inferences show a substantial dependence on the tree-building method used. In the case of Afro-Asiatic, for non-feature trees, VSO is the most probable ancestor for lower  $t$  values, but as  $t$  increases, the posterior probability of VSO decreases and those of SOV and SVO rise. One of the two subject-initial orders gains probability faster than the other: With the geographic and combination trees, SOV gains probability faster and eventually becomes strongly preferred over SVO. With the genetic trees, however, SVO outpaces SOV instead, although it never becomes substantially more probable than SOV.

In Indo-European, SOV is clearly a more probable ancestor than SVO for the geographic, feature, and combination trees, while SVO is preferred over SOV only in the genetic case.

Nilo-Saharan shows the most complex behavior, within and between tree-generation methods. VSO has a nonnegligible probability of being the ancestral word order for most  $t$  values for all four tree methods. It competes with SOV and SVO in all trees except for feature, where it competes only with SVO. The details of this competition vary considerably. We attribute this unusual behavior to Nilo-Saharan being the only family in our sample that has SOV, SVO, and VSO all present in substantial proportions.

Whereas our inferences about word-order stability and preferred directions of change showed quite good agreement across the different tree-generation methods, many of our inferences about word-order ancestry exhibit quite strong dependence on the method used. The situation is not too dire, however. If we are willing to ignore the feature trees, which seems reasonable given their extremely poor fit to the data, then four of our seven families show essentially method-independent results. Afro-Asiatic, Indo-European, and Nilo-Saharan are the only families

with substantial variation across features. This is presumably due to the fact that the other four families in our sample are strongly dominated by a single word order, and as such, their inferred ancestries are much less dependent upon the particulars of their tree topology.

**Conclusion.** The BIC values in Table S2 make it clear that the combination tree-generating method that underlies the results presented in the main text provides the best model for our purposes, out of the alternatives we have described. Nevertheless, it is important to assess to what extent our findings are dependent upon this method. The results listed above suggest that our conclusions about word-order stability and preferred directions of change are almost completely independent of the tree-generation method used. Our conclusions about ancestral word orders show more variation, although there is broad agreement for five out of the seven families if we disregard the feature method.

**Fitting Individual  $Q$  Matrices to Each Family.** An analysis where six language families share a common  $Q$  matrix was chosen for our primary results. An alternative is to allow each family to have its own distinct  $Q$  fitted to it. These two approaches are opposite extremes of a continuum, with intermediary points representing different amounts of family dependence in the change dynamics.

Word-order change can happen for a variety of reasons. Some of these presumably have a degree of randomness about them, such as social selection (where linguistic idiosyncrasies of individuals held in high social standing are emulated and become population-wide features) or change due to borrowing from neighboring, possibly genetically unrelated (or only very distantly related) languages. Word-order changes that happen in one language family for these reasons should not necessarily be expected to happen with a similar frequency in other families. Therefore, some degree of family dependence in change dynamics is to be expected. However, if this sort of undirected change accounted for the majority of word-order change, we should expect to observe a relatively uniform cross-linguistic distribution. The fact that we find such striking nonuniformity suggests that there is a degree of family-independent structure to the word-order change dynamics. This structure presumably comes from functional concerns.

At one end of our continuum, fitting one  $Q$  matrix to all seven families attempts to capture this underlying functionally driven change, treating the various family-dependent components of change as noise. At the other end, fitting six different  $Q$  matrices to the different families has less capacity to separate functionally from nonfunctionally driven change. Each of the six  $Q$  matrices is also fit from less data; therefore it is more difficult to distinguish rare changes from common changes. We therefore consider the single  $Q$  approach used in the main text to offer the most insight. For completeness, we consider the case of individual  $Q$  matrices here. We use the combination tree-generation method throughout.

**Word-order dynamics.** When one  $Q$  matrix is shared between all seven families, we see that SOV and SVO typically have very similar posterior probabilities of being the most stable word order. A stark contrast is observed when we fit one  $Q$  matrix to each family. Each family has one word order with a much higher probability of being the most stable order than the others. These are: SOV for Afro-Asiatic (0.84), SVO for Austronesian (0.69), SOV for Indo-European (0.73), SVO for Niger-Congo (0.99), SVO for Nilo-Saharan (0.78), SOV for Sino-Tibetan (1.00), and SOV for Trans-New Guinea (1.00). In every case except Afro-Asiatic, the word order that is most common in each family is the most stable word order for the majority of the MCMC samples. The most homogeneous families show the least variation across samples, while those families where several word orders are present in substantial numbers occasionally produce samples where nonmajority word orders are the most stable. On



the whole, the results are reasonably well approximated by saying that each family's posterior distribution over  $Q$  matrices assigns the six word orders probabilities of being the most stable order roughly proportional to their frequency in that language.

**Ancestral word orders.** Fig. S11 shows the posterior distributions over ancestral word orders for the seven families, where each family has its own  $Q$ . Broadly speaking, those families that are dominated by one word order have that word order being the only word order with nonnegligible posterior probability of being the ancestral word order for the majority of  $t$  values. Less heterogeneous families typically assign both SOV and SVO nonnegligible posterior probability for most values of  $t$ . These results are not especially surprising given our observations regarding the posterior distributions over  $Q$  matrices.

These plots feature notably more ambiguity with regard to ancestral word order than the majority of the common- $Q$  plots, and this is due to the underconstrained nature of the family-specific  $Q$  matrices. The clearest example of this is Indo-European, which is dominated by SOV and SVO languages, which occur with roughly equal frequency. The only way that either SOV or SVO could be strongly preferred over the other as an ancestor is if we have strong evidence that these orders differ in their stability, or that change toward one is more common than toward the other. If  $Q$  is fit only to Indo-European data, it is relatively unlikely that it will include any such asymmetries in the word-order change dynamics. However, if  $Q$  is fit to all seven families, many of which do not feature SOV and SVO in roughly equal numbers, then we can infer these differences in dynamics

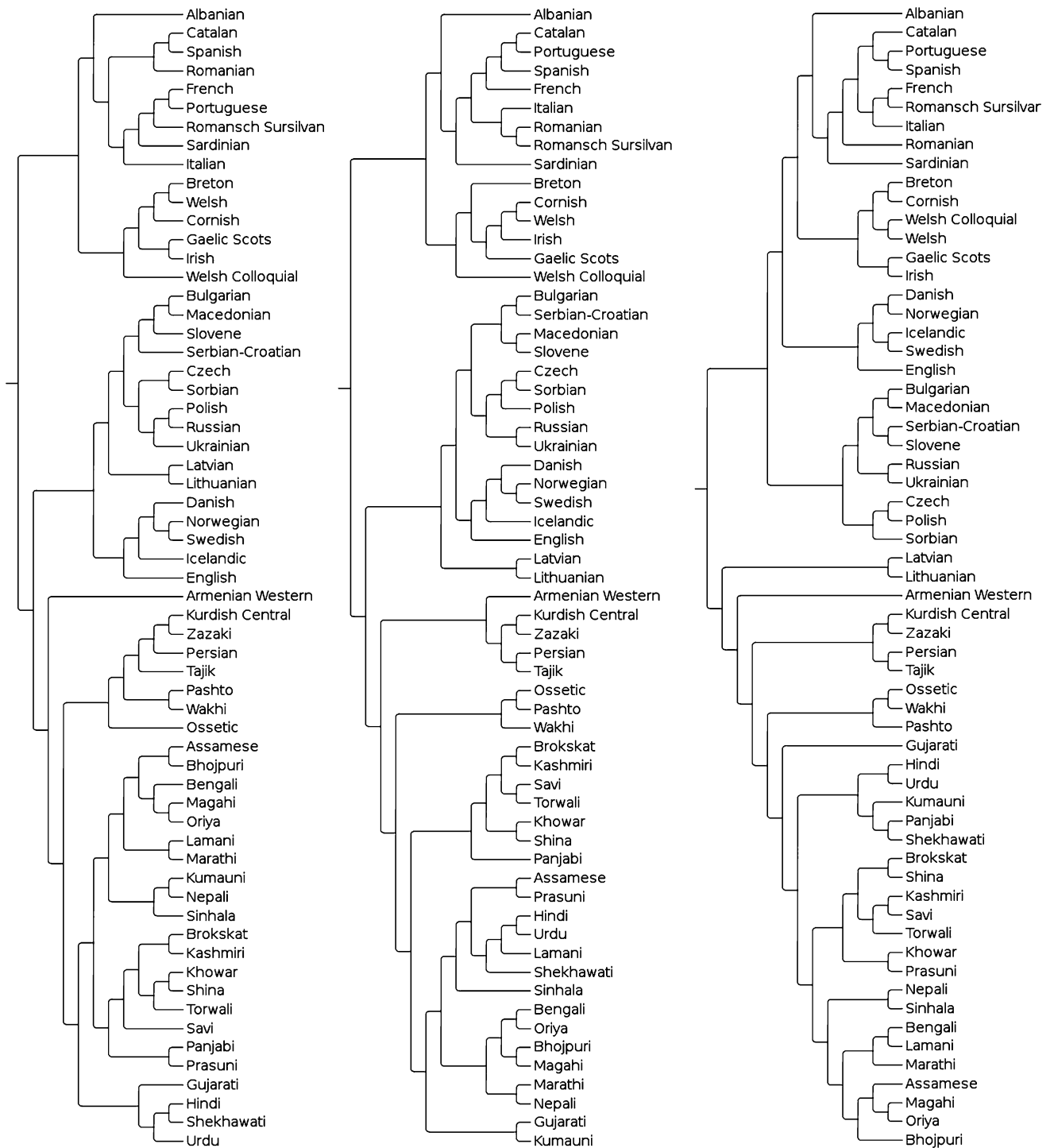
and thereby resolve the ambiguous ancestry. Thus, Fig. 2 shows a strong preference for SOV ancestry in Indo-European, while Fig. S11 shows SOV and SVO having roughly equal posterior probability of being ancestral.

**Conclusion.** The inferred  $Q$  matrices found for each family on its own make it clear that this approach leads to widespread overfitting. Each family's  $Q$  matrix makes that family's common word orders stable and its rare word orders unstable. This leads to a good explanation of the data, as indicated by the higher likelihood values seen in Table S2. However, as the BIC values in Table S2 indicate, this improved fit is not proportional to the drastically increased model complexity. The posterior distributions over ancestral word orders shown in Fig. S11 also show that this approach has far less power to resolve ancestral word orders. We believe it is clear that the approach used in the main text, of fitting a common- $Q$  matrix to all seven families, is the better of the two approaches.

None of this suggests that variation in word-order dynamics across families does not exist. A promising direction for future research would be to develop a model that permits constrained variation in dynamics across families. Such a model might endeavor to infer a  $Q$  matrix corresponding to the influence of family-independent factors (nonlinguistic cognition, cultural transmission dynamics), with each family's history modeled using a  $Q$  that is the sum of this family-independent  $Q$  and a small matrix  $Q_{\text{FAM}}$  of family-dependent variations that can account for family-specific influences on change, such as language contact.

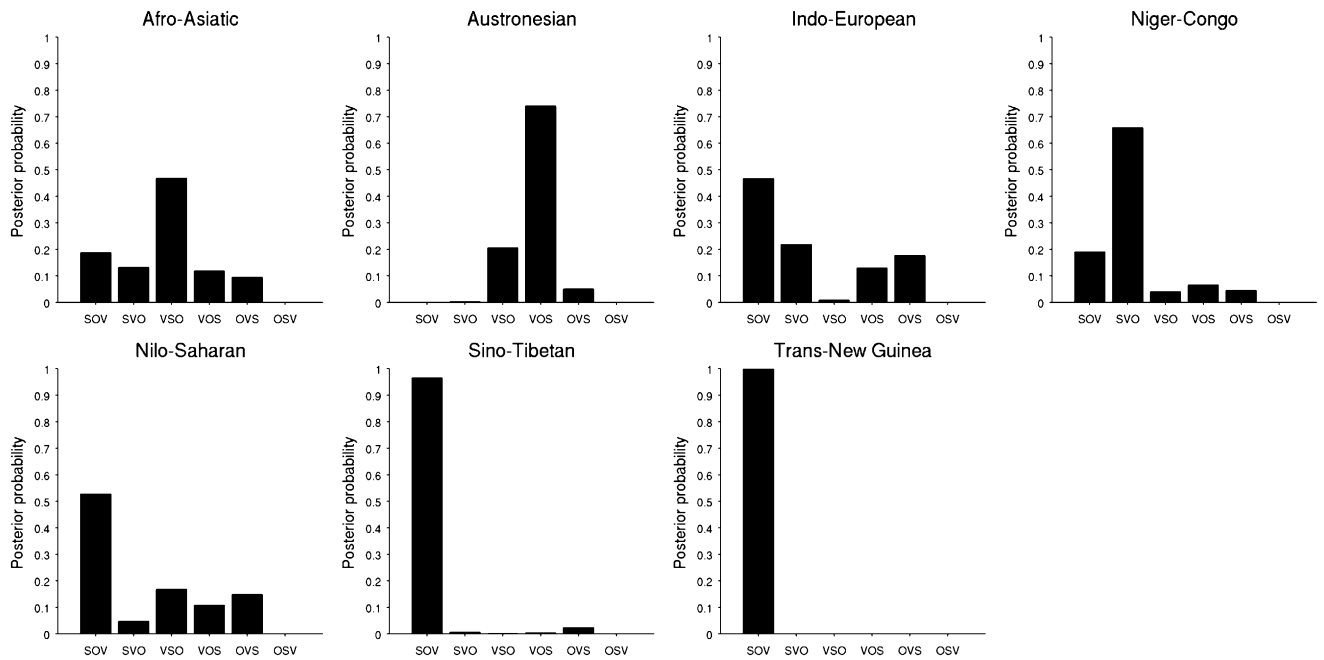
1. Dryer M, Haspelmath M, eds (2011) *The World Atlas of Language Structures Online* (Max Planck Digital Library, Munich).
2. Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6(2):461–464.

3. Pagel M (2009) Human language as a culturally transmitted replicator. *Nat Rev Genet* 10(6):405–415.

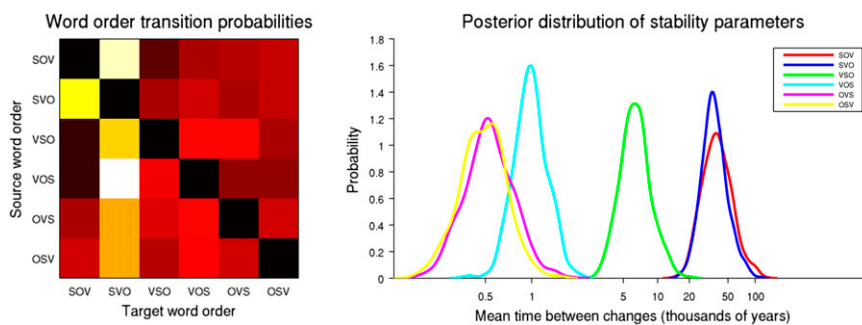


**Fig. S1.** Three Indo-European family trees, generated by applying neighbor joining to randomly generated matrices built using the combination method. The trees display some variation (e.g., French has a different nearest neighbor in each of the trees), but the variation is constrained by data (the language most closely related to French is always another Romance language). None of the trees is exactly correct according to modern Indo-European linguistics, but a large number of trees generated by this method will constitute an exploration of the space of trees characterized by sensible variation around the ground truth. Any statements about basic word order that are strongly implied by the majority of the trees in such a large sample are likely to be true of the correct tree as well (which for some families is not known with confidence).

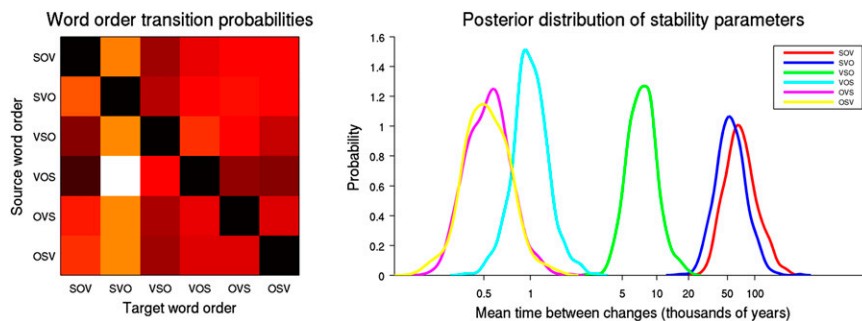




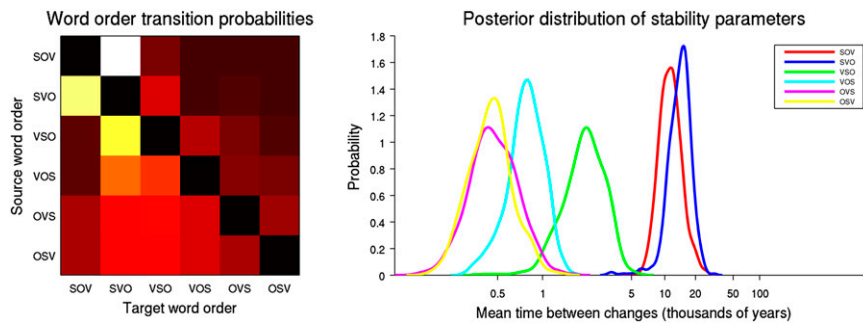
**Fig. 54.** Posterior probability of different ancestral word orders for each language family, using the present-day distribution as a prior. SOV is the most probable ancestor for four of the seven families.



**Fig. 55.** Results of inferring a single mutation matrix  $Q$  for all six language families, using the geographic tree method. (Left) Heat map showing the transition probabilities between word orders. Higher intensity indicates more probable transitions. (Right) Inferred posterior distributions of stability parameters for each word order. Horizontal axis shows mean time between transitions.

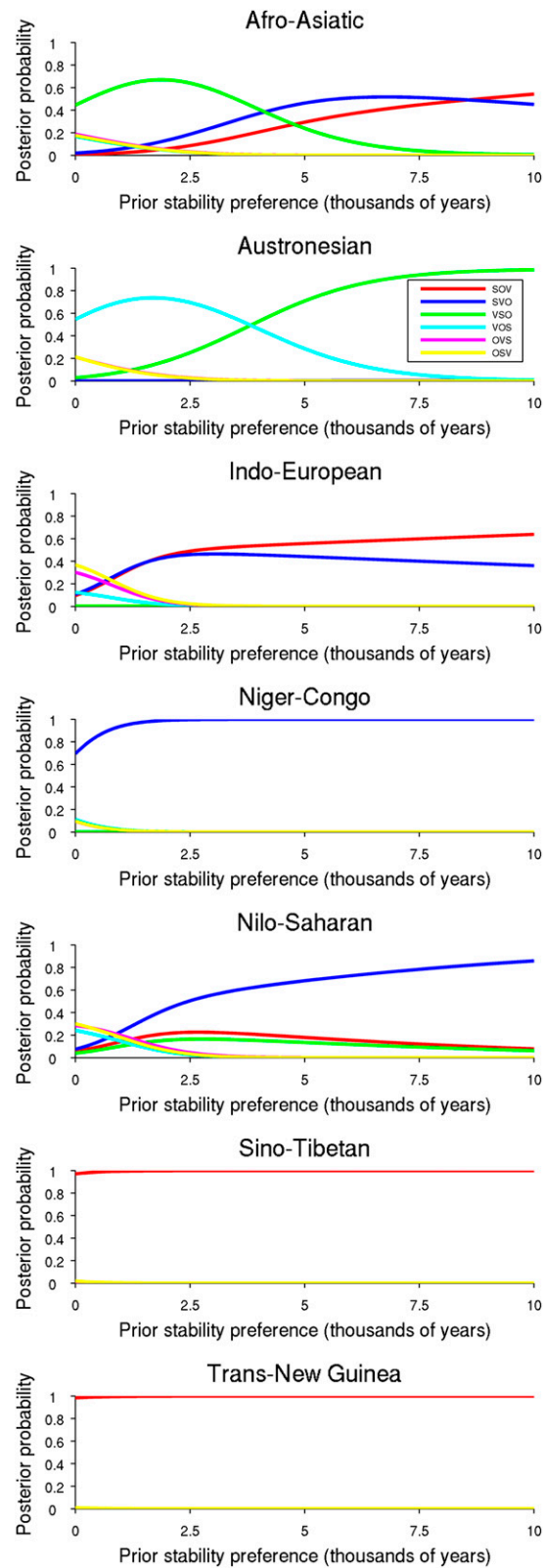


**Fig. 56.** Results of inferring a single mutation matrix  $Q$  for all six language families, using the genetic tree method. (Left) Heat map showing the transition probabilities between word orders. Higher intensity indicates more probable transitions. (Right) Inferred posterior distributions of stability parameters for each word order. Horizontal axis shows mean time between transitions.

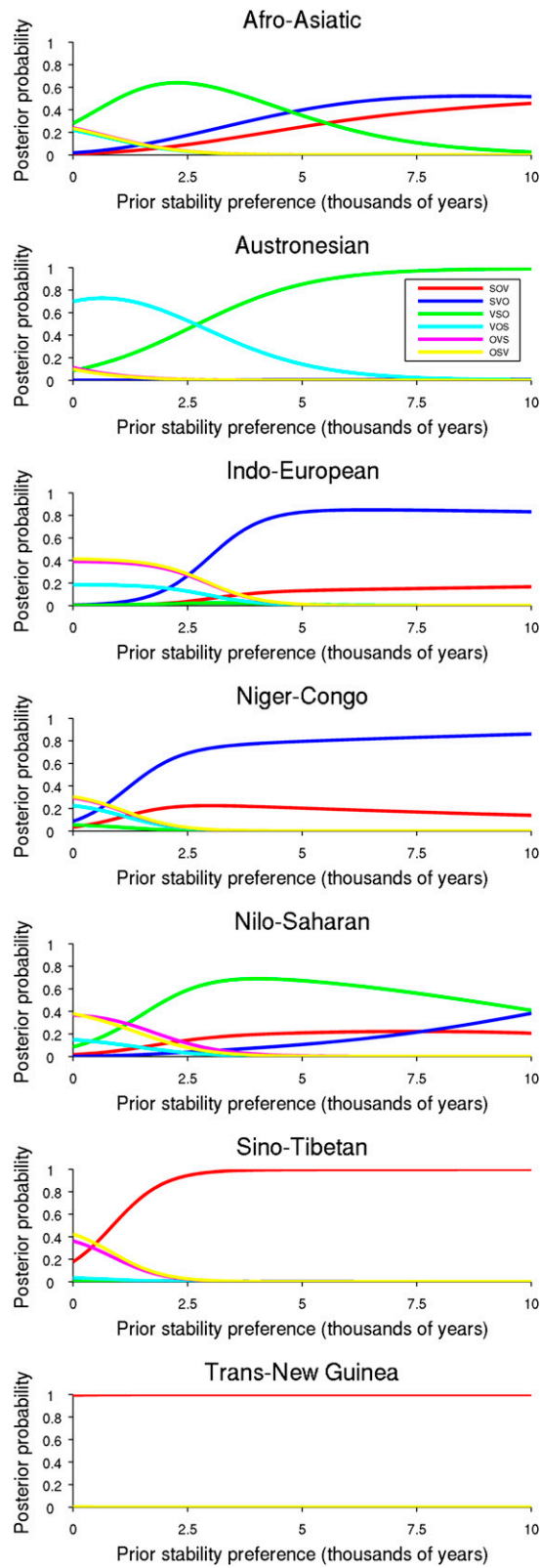


**Fig. S7.** Results of inferring a single mutation matrix  $Q$  for all six language families, using the feature tree method. (*Left*) Heat map showing the transition probabilities between word orders. Higher intensity indicates more probable transitions. (*Right*) Inferred posterior distributions of stability parameters for each word order. Horizontal axis shows mean time between transitions.

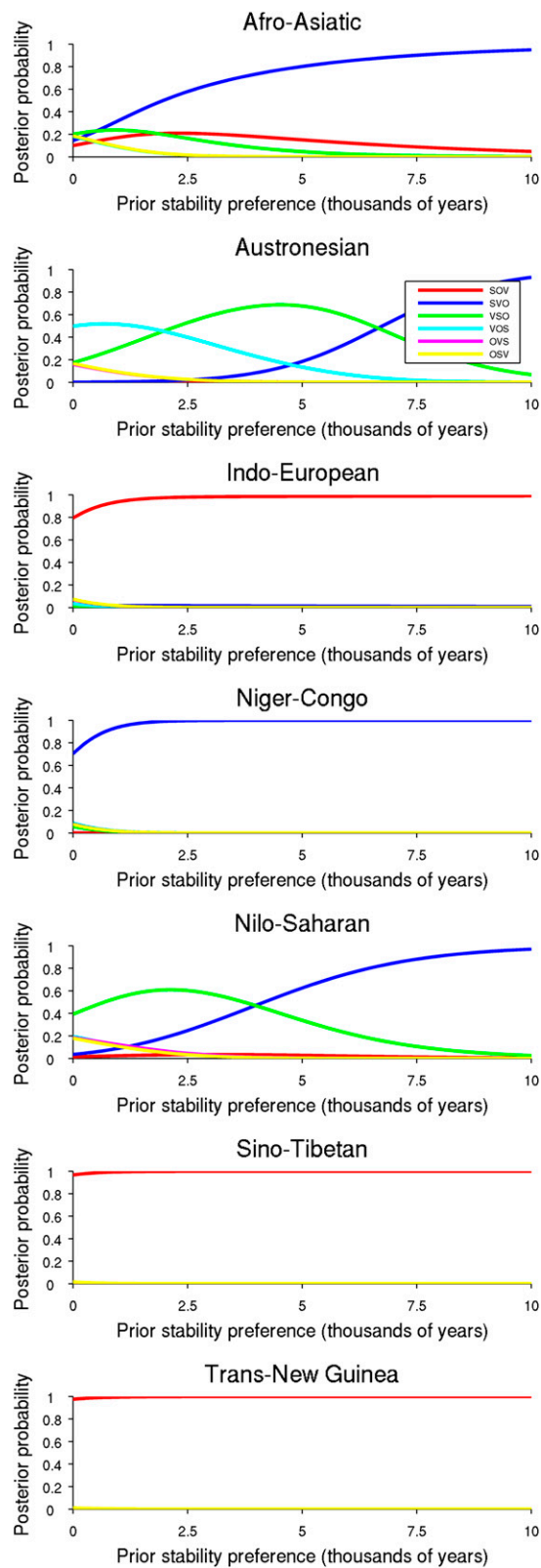




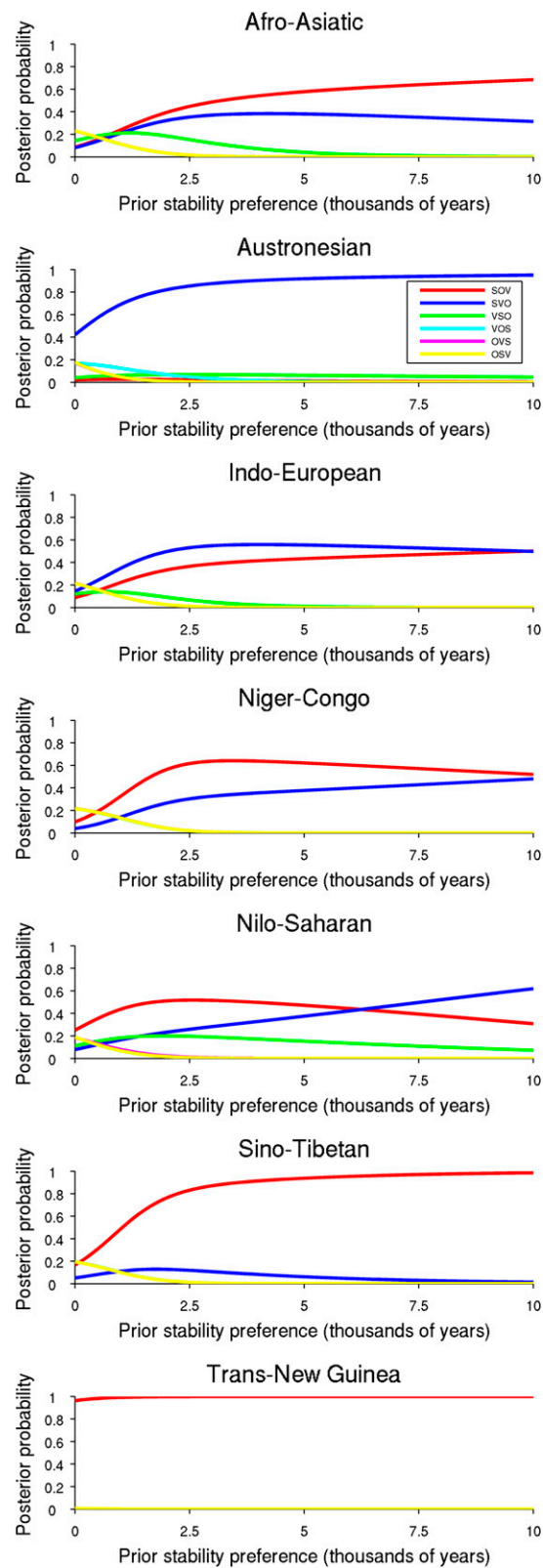
**Fig. S8.** Posterior probabilities of different ancestral word orders for each language family, using the geographic tree method and the same parameterized family of priors from the main text. Each prior assigns word orders probabilities proportional to the probability of a language starting with that word order not changing after a certain length of time. The longer the timespan, the stronger the preference for stable ancestral word orders.



**Fig. S9.** Posterior probabilities of different ancestral word orders for each language family, using the genetic tree method and the same parameterized family of priors from the main text. Each prior assigns word orders probabilities proportional to the probability of a language starting with that word order not changing after a certain length of time. The longer the timespan, the stronger the preference for stable ancestral word orders.



**Fig. S10.** Posterior probabilities of different ancestral word orders for each language family, using the feature tree method and the same parameterized family of priors from the main text. Each prior assigns word orders probabilities proportional to the probability of a language starting with that word order not changing after a certain length of time. The longer the timespan, the stronger the preference for stable ancestral word orders.



**Fig. S11.** Posterior probabilities of different ancestral word orders for each language family, using the combination tree method and where each family has had a separate Q matrix fitted to its data. Each prior assigns word orders probabilities proportional to the probability of a language starting with that word order not changing after a certain length of time. The longer the timespan, the stronger the preference for stable ancestral word orders.

**Table S1. Number of languages having each of the six possible basic word orders for each of the seven language families used in the present study**

Language family	SOV	SVO	VSO	VOS	OVS	OSV
Afro-Asiatic	21	33	8	0	0	0
Austronesian	14	95	25	10	1	1
Indo-European	30	27	4	0	0	0
Niger-Congo	22	175	0	0	0	0
Nilo-Saharan	16	30	13	0	1	0
Sino-Tibetan	81	8	0	0	0	0
Trans-New Guinea	56	0	0	0	0	0
Total	240	368	50	10	2	1

**Table S2. Log-likelihoods and BIC values for all combinations of tree building method and Q matrix counts**

Tree method	Q matrix	Log-likelihood	BIC
Geographic	common	-306.32	843.82
Genetic	common	-287.89	806.96
Feature	common	-777.67	1786.52
Combination	common	-270.82	772.82
Geographic	individual	-273.64	1934.35
Genetic	individual	-258.65	1910.37
Feature	individual	-564.96	2516.99
Combination	individual	-243.28	1873.61

Regardless of whether  $Q$  is common to all languages or each language gets its individual  $Q$ , the combination method results in the highest log-likelihood and lowest (best) BIC. Regardless of which tree method is used, the common- $Q$  approach yields the lowest (best) BIC, due to the drastic increase in the number of free parameters (36 for common, 216 for individual).