John Benjamins Publishing Company

Shedding more light on language classification using basic vocabularies and phylogenetic methods

A case study of Uralic*

Kaj Syrjänen[a], Terhi Honkola[b], Kalle Korhonen[c], Jyri Lehtinen[c], Outi Vesakoski[b] and Niklas Wahlberg[b]
[a]University of Tampere / [b]University of Turku / [c]University of Helsinki

Encouraged by ongoing discussion of the classification of the Uralic languages, we investigate the family quantitatively using Bayesian phylogenetics and basic vocabulary from seventeen languages. To estimate the heterogeneity within this family and the robustness of its subgroupings, we analyse ten divergent sets of basic vocabulary, including basic vocabulary lists from the literature, lists that exclude borrowing-susceptible meanings, lists with varying degrees of borrowing-susceptible meanings and a list combining all of the examined items. The results show that the Uralic phylogeny has a fairly robust shape from the perspective of basic vocabulary, and is not dramatically altered by borrowing-susceptible meanings. The results differ to some extent from the 'standard paradigm' classification of these languages, such as the lack of firm evidence for Finno-Permian.

Keywords: historical linguistics, computational phylogenetics, Uralic languages, basic vocabulary

1. Introduction

Computational methods originally developed for the classification of biological species have gradually seeped into linguistics and today are widely used in

* We thank Tapani Salminen, Antti Leino, Urho Määttä, as well as several anonymous commentators and reviewers for their helpful comments on the manuscript at its various stages. We would like to thank Sami Merilaita, Dan Dediu and Mark Pagel for valuable discussion. The lexical dataset was compiled with the assistance of the Institute for the Languages of Finland (Kotimaisten kielten

ISSN 0176–4225 / E-ISSN 1569–9714

answering questions about language history, classification and typology (e.g. Gray & Jordan 2000, Ringe et al. 2002, Gray & Atkinson 2003, McMahon & McMahon 2003, 2005, Atkinson et al. 2005, Nakhleh et al. 2005, Dunn et al. 2008, Greenhill & Gray 2009, Bowern 2010, Lee & Hasegawa 2011). To some extent, these computational methods continue the lexicostatistical tradition, and in fact, there are various connections in the evolution of these tool sets.[1] Increasingly, lexicostatistical tools are being supplanted by phylogenetic software.[2]

In this paper we apply a quantitative method imported from biology — Bayesian phylogenetic analysis — to examine the classification of the Uralic languages. Despite extensive research for well over a century, the hierarchy and subgroupings of this language family, particularly the early branchings such as Samoyed, Permian and Ugric, remain topics of lively discussion and research. The discussion has led to various proposals for classifying Uralic languages that clash, at least in part, with the traditional textbook classification (e.g. Häkkinen 1984, Michalove 2002, Salminen 2002, Kulonen 2002, Häkkinen 2009, Saarikivi 2011). Most Uralic research remains non-quantitative, and classification is rarely placed on an explicit scale of certainty.[3] Our main objective is to address the issue with the help of Bayesian phylogenetics, which provides tools for both reconstructing the phylogeny and estimating the level of certainty and evidential confidence for the subgroupings, giving us further insight into Uralic classification.

keskus). The authors also thank the following for checking the dataset: Marja Torikka (Karelian), Santra Jantunen (Veps and Livonian), Eve Mikone (Estonian), Kaarina Vuolab-Lohi (North Saami), Eino Koponen (Ume Saami and Skolt Saami), Natalia Chinaeva (Erzya), Arto Moisio (Meadow Mari), Svetlana Lumme (Komi), Olga Titova (Udmurt), Judit Varga and Zoltán Balogh (Hungarian), Susanna Virtanen (Northern Mansi), Merja Salo (Eastern Khanty), Tapani Salminen (Tundra Nenets and Selkup). We are grateful to Ilpo Tammi for producing the Uralic language map, and to Professor Glenda D. Goss for proofreading this manuscript. The authors are responsible for all remaining mistakes. The authors acknowledge funding from Kone Foundation, the Finnish Academy of Science and Letters, the Academy of Finland and the University of Helsinki.

1. For example, various improvements in lexicostatistic methodology proposed by Sankoff in the 1970s — such as the introduction of the gamma distribution (Sankoff 1973) — are connected to innovations in quantitative biology around the same time (e.g. Embleton 1986, Atkinson & Gray 2005).

2. A significant trend encouraging the adoption of modern methods from evolutionary biology is the growing interest in parallels between biological evolution and language change (e.g. Croft 2000, 2006) and, on a wider scale, parallels between biological and cultural evolution. The influence of this trend has been noted by, e.g., Croft (2006, 2008) and Pagel (2009).

3. A noteworthy exception to this is a recent quantitative study conducted in parallel with this one (Honkola et al. 2013), which focused on the divergence time estimates for the Uralic languages.

The analyses examine basic vocabulary using MrBayes (Huelsenbeck & Ronquist 2001, Ronquist & Huelsenbeck 2003), which yields a distribution of phylogenetic trees. To obtain a better view of how heterogeneously the languages can be classified and to see the impact of the quantity and quality of the data on the results, the analyses are repeated on several quantitatively and qualitatively distinct subsets, including three different basic vocabularies (a 100-item Swadesh list, a 207-item Swadesh list and the Leipzig-Jakarta list), a subset of all meanings from the lists that are not documented as borrowings, five subsets of meanings with documented borrowings and a list combining all of the aforementioned meanings together.

In §2 we set the stage by introducing the Uralic language family, highlighting some of the outstanding questions about its classification. After this we discuss the qualitative and quantitative aspects of basic vocabularies in §3, provide an overview of the dataset in §4 and in §5 introduce how Bayesian phylogenetic analysis operates, how its results are interpreted and what settings were used in the analyses. Finally, we present and discuss the results in §6 and §7 and conclude the paper in §8.

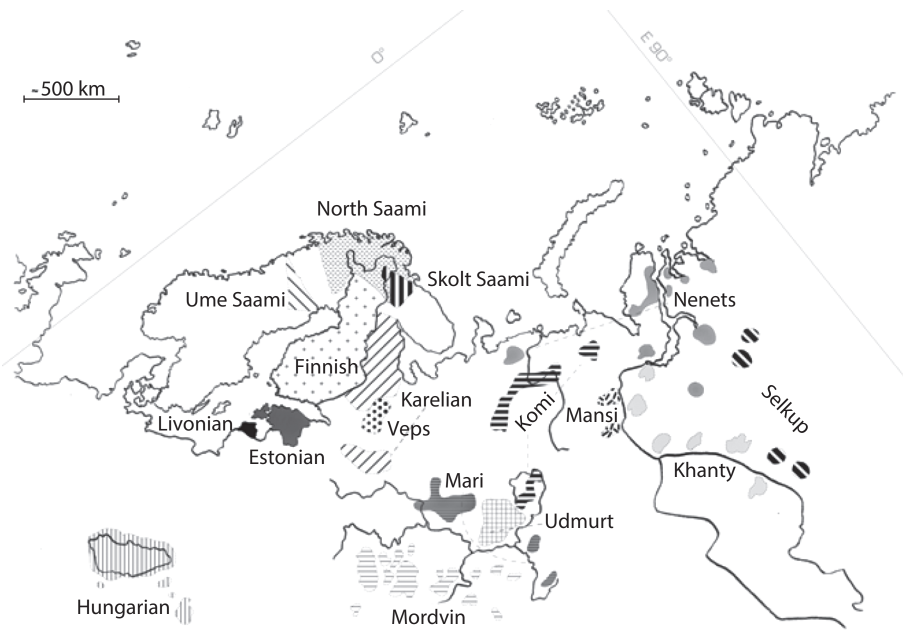


Figure 1. Map of the Uralic languages showing the locations of the languages under investigation. Geographical areas are based on information from Abondolo (1998).

2. The Uralic language family

The Uralic language family consists of 47 languages spoken by about 25 million speakers scattered across northeastern Europe and Siberia (see Salminen 2007, Janhunen 2009). Relationships among Uralic languages have been actively studied from the beginnings of modern historical comparative linguistics and to some extent before (see Hovdhaugen et al. 2000), mainly with traditional, non-quantitative methods (although see Raun 1956, Taagepera 1994, Michalove 2002, Tambovtsev 2004, Künnap & Taagepera 2005). Figure 1 shows the geographical distribution of the languages examined, Table 1 lists the Uralic languages in their main groupings and Table 2 lists the traditional intermediate levels.

Table 1. The Uralic languages in their main subgroupings. Extinct languages are marked with †.

Grouping	Languages
Saami	South Saami, Ume Saami, Pite Saami, Lule Saami, North Saami, Inari Saami, Kemi Saami†, Skolt Saami, Akkala Saami†, Kildin Saami, Ter Saami
Finnic	Finnish, Ingrian, Karelian, Veps, Vote, Estonian, Livonian†, Võro-Seto, Olonetsian, Lude
Mordvin	Erzya, Moksha
Mari	Meadow Mari, Hill Mari
Permian	Komi, Permyak, Udmurt
Samoyed	Nganasan, Tundra Enets, Forest Enets, Yurats†, Tundra Nenets, Forest Nenets, Northern Selkup, Central Selkup, Southern Selkup, Kamas†, Mator†
Mansi	Northern Mansi, Eastern Mansi, Western Mansi†, Southern Mansi†
Khanty	Northern Khanty, Southern Khanty†, Eastern Khanty
–	Hungarian

Table 2. The traditional intermediary subgroupings of Uralic languages.

Subgrouping	Groupings / languages
Uralic	Finno-Ugric, Samoyed
Finno-Ugric	Ugric, Finno-Permian
Finno-Permian	Permian, Finno-Volgaic
Finno-Volgaic	Mari, Mordvin, Finno-Saami
Finno-Saami	Finnic, Saami
Ugric	Hungarian, Ob-Ugric
Ob-Ugric	Mansi, Khanty

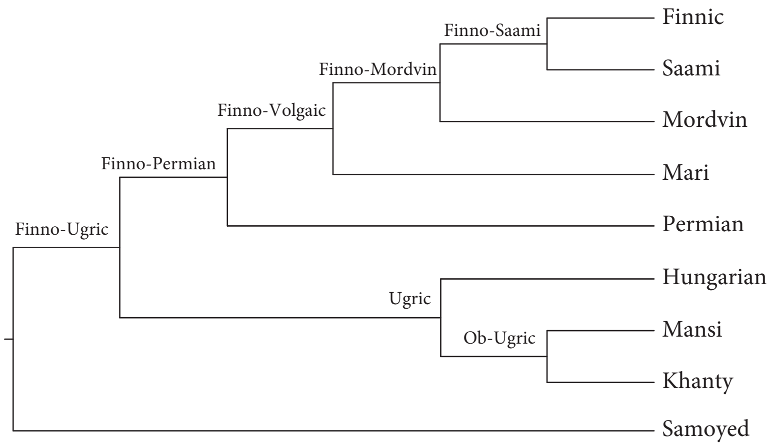


Figure 2. The classification of Uralic languages, based on Korhonen (1981).

The traditional view is that the Uralic languages descend from a proto-language in a fairly tree-like pattern. The Finno-Ugric *Stammbaum* proposed by Otto Donner in the late nineteenth century formed the foundation for what is often called the 'standard paradigm' of Uralic phylogeny (Hovdhaugen et al. 2000: 178–179). In Donner's interpretation, the ancestral Finno-Ugric language split into a Ugric branch and a Finno-Permian one, the latter then separated into Permian and Finno-Volgaic branches; finally, after the divergence of Mari and Mordvin,[4] the remaining Finno-Saami group split into Saami and Finnic. With the inclusion of Samoyed alongside Donner's Finno-Ugric classification, the tree becomes the textbook phylogeny for the Uralic language family, with Samoyed as the first language to diverge. Figure 2 simplifies Korhonen's (1981) Uralic tree, reflecting the textbook classification fairly closely.

Although it is generally accepted that these languages form a genealogical whole, the textbook classification has been debated since its inception. This is not surprising, considering that linguistic material from these languages only dates back some 1,000 years, with the bulk of the older language material originating from nineteenth century fieldwork (Sinor 1988). Most debate is focused on early branchings, including:

a. <u>Ugric</u>. Michalove (2002) and others raise doubts about the validity of Ugric, while accepting the Ob-Ugric clade as valid. With respect to typological

4. In the early classifications, such as the one proposed by Donner, Mari and Mordvin formed a Volgaic subgrouping. However, Mordvin and Mari are currently considered separate branches because Volgaic is not supported by shared phonological or morphological or lexical innovations (Itkonen 1997, Michalove 2002). For classifications that include Volgaic, see Ruhlen (1987).

compactness, Tambovtsev (2004) has likewise commented on the apparent artificiality of Ugric. Salminen (2002) has suggested that rather than being a 'true' genealogical subgrouping, Ugric should be regarded as reflecting a combination of genetic and areal factors. We note Janhunen's (2000) unusual Ugric subgrouping in which Hungarian and Mansi are more closely linked, with Khanty on the outside and, consequently, the Ob-Ugric clade absent.[5] Most of these challenges to Ugric or Ob-Ugric unity are based on the apparent difficulty in reconstructing sound patterns for (Ob-)Ugric proto-forms.

b. Samoyed, in particular its traditional position as the first clade to diverge from Proto-Uralic. Although the traditional placement has been believed to stand on firm lexical ground (e.g. Janhunen 2000, Salminen 2002), this has been challenged on phonological and grammatical grounds, on the basis of which Samoyed could also share the historical stage with Ugric or Ob-Ugric (e.g. Salminen 2002, Häkkinen 2009).

c. Permian, in particular its traditional position. Based on sound correspondences, Häkkinen (2009) has proposed a possible closer association of Mari and Permian. In addition Salminen (2002) notes that the connection between Mari and Permian likely reflects a combination of areal and genetic factors.

Uralic phylogeny has also received considerable attention in the last few decades, and several classifications contrast with the traditional family tree by being more polytomous, or 'bush-like', thus highlighting points where the bifurcating Donneresque tree could be considered misleading. Highly polytomous classifications have been given, for example, in Häkkinen (1983), Salminen (1999, 2007) and Saarikivi (2011). All have little internal hierarchy and show the Uralic language branching directly into multiple branches instead of splitting binarily into Finno-Ugric and Samoyed. Less dramatic variants include Kulonen's 2002 classification, which gives more certainty to some branches, and splits the Uralic root three ways — into Finno-Permian, Samoyed and Ugric. Figures 3 and 4 (adapted from Salminen (2007) and Kulonen (2002) respectively) show the differences between these perspectives and the more traditional, bifurcating Uralic phylogeny shown in Figure 2. General references still tend to prefer a more conservative, Donneresque diagram, in which most of the intermediate levels remain intact (e.g. Crystal 1987: 304, Ruhlen 1987: 68, Matthews & Polinsky 1997, Price 1998: 482, Austin 2008).

For the most part, the uncertain positions of languages are fairly well known qualitatively; the quantitiative character of these uncertainties, however, has not

5. This has been associated with the areal distribution of the correspondences shared by Mansi, Khanty and Hungarian (e.g. Itkonen 1997).

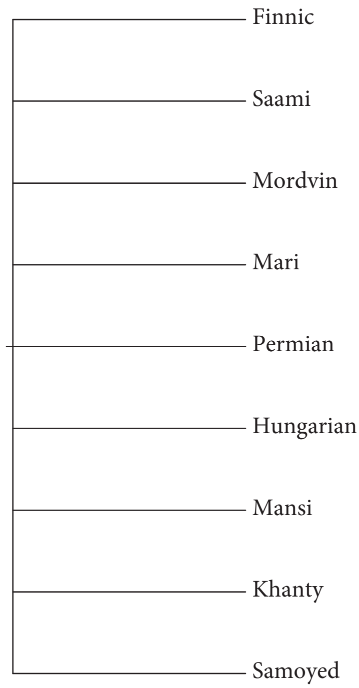


Figure 3. The classification of the Uralic languages, based on Salminen (2007).

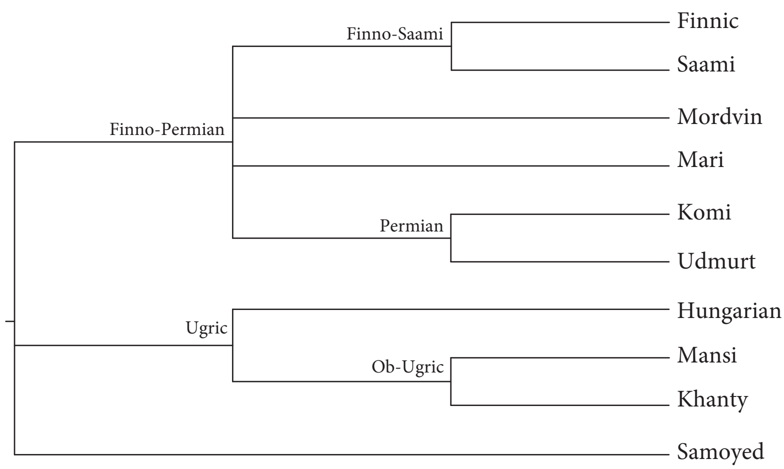


Figure 4. The classification of the Uralic languages, based on Kulonen (2002).

been researched to the same extent. The approach taken here seeks balance by giving numerical estimates for the ambiguity and the level of evidential confidence for the subgroupings at the level of basic vocabulary.

3. Quantitative and qualitative considerations of basic vocabulary lists

The present analyses revolve around basic vocabulary, concepts denoted by words that are relatively stable over the history of a language, universally present in different languages, morphologically and semantically simple and resistant to borrowing. Basic vocabulary lists are the staple of quantitative historical linguistics and have been extensively used since the beginnings of lexicostatistics in the 1950s. Standardized sets of basic vocabulary have been compiled for this purpose. Perhaps the most popular of these are the 200-item Swadesh list (Swadesh 1952) and its later 100-word counterpart (Swadesh 1955), from which various problematic items (culture-specific meanings, meanings without reliable matches in all languages, ambiguous object words, non-independent meanings that may cause duplication) were eliminated in order to improve the quality of the list. Alternatives to the Swadesh lists include Tadmor's (2009) Leipzig-Jakarta list, for which the meanings were chosen on the basis of several quantitatively evaluated criteria (susceptibility to borrowing, historical age in the language family, morphological simplicity and representativeness of the meaning in the vocabularies of different languages) assessed from languages around the world.

Both the quantitative and the qualitative aspects of the dataset employed have an effect on how languages are classified by computational methods. The effect of quantity is well illustrated in Embleton's (1986: 89–93) comparisons of a 100-item dataset, a 200-item dataset and a 500-item dataset on simulated language data. She concluded that the more data one has, the more accurate the classification is. Notably, however, her analyses also showed that the improvement in accuracy is more significant when the list is expanded from 100 items to 200 items than when it is expanded from 200 items to 500 items. The 200-item list could thus be seen as a fair compromise between practical considerations and overall accuracy. Her analyses also highlight one problem in using the newer basic vocabulary lists such as Swadesh100 or Leipzig-Jakarta, the smaller number of items.

However, quality is also important for basic vocabularies. Different meanings do not generally change at a homogeneous rate, with word frequency being one factor affecting the rate of change (e.g. Dyen et al. 1967, McMahon & McMahon 2003: 40–41, Pagel et al. 2007). Qualitative factors such as borrowings within basic vocabulary meanings also affect different language families in unique ways. McMahon & McMahon (2003) have illustrated the significance of quality with Indo-European data and bootstrapping analyses, showing that even with 200 items we can still see considerable heterogeneity in the contributions that the different meanings make to the accuracy of the classification. McMahon & McMahon (2005: 94–96) have also pointed out the generally smaller proportion of loanwords in the 100-word Swadesh list in comparison with the 200-item Swadesh list by examining languages with high levels of borrowing from a dataset compiled by Kessler (2001). In this context McMahon & McMahon (2005) suggested that a borrowing level of 10% is enough to pose problems for quantitative classification. However, more recently Greenhill et al. (2009) estimated that in the case of non-systematic borrowing,[6] borrowing rates of up to 20% would have a very small effect on the shape of the tree produced by Bayesian methods. This difference, as they point out, may be due to methodological differences (they used character-based methods, unlike McMahon & McMahon, who used distance-based methods), but also because McMahon & McMahon's estimate was based on cases of systematic borrowing. While the results from Greenhill et al. suggest that the quantitatively better 200-item Swadesh list would not pose a major problem for some methods, the smaller proportion of loanwords generally gives some grounds for favouring more qualitative lists such as the 100-item Swadesh list or the Leipzig-Jakarta list, or modifying a standardized list to be more suitable for a specific set of languages, as is often done.[7]

In summation, it remains difficult to specify the best balance between quantity and quality for basic vocabulary analyses, as these two factors appear to have different effects on different kinds of analyses. Consequently, in our investigation we divided the basic vocabulary data into subsets that included different standardized basic vocabulary lists, quantitative differences (different numbers of meanings) and qualitative differences (different levels of borrowing-susceptibility), and conducted identical analyses for each of these subsets. This gave us insight into how the various qualitative and quantitative differences are reflected in the results, helping us obtain a better idea of how much overall variation exists in the borrowing patterns and the vertical inheritance patterns of the Uralic basic vocabulary items.

4. The dataset of Uralic languages

The data come from seventeen Uralic languages, chosen to provide good coverage of traditional Uralic subgroupings. The languages are: Finnish, Karelian, Veps,

6. Greenhill et al. (2009: 2304) refer to sporadic borrowing events that do not significantly affect the linguistic system as a whole as 'non-systematic borrowing', noting that it is essentially the same as Bloomfield's (1933) 'cultural borrowing'. By contrast, in 'systematic borrowing' traits flow predominantly from one culture to another, introducing systematic biases in the data.

7. McMahon & McMahon (2005: 33; 156) exemplify modified basic vocabularies, including the CALMSEA list (Matisoff 2000), the Australian language lists by O'Grady (1960) and Alpher & Nash (1999) and Heggarty's CALMA list (McMahon et al. 2005). The Bantu basic vocabulary by Bastin et al. (1999) is another noteworthy example.

Estonian, Livonian, North Saami, Ume Saami, Skolt Saami, Erzya, Meadow Mari, Komi, Udmurt, Hungarian, Northern Mansi, Eastern Khanty, Tundra Nenets and Selkup (see Figure 1 for the geographical distribution of these languages). Dictionaries (common and etymological) were the primary sources for compiling the dataset, and only languages with sufficient reference material were included. Each language was double-checked by at least one specialist in or native speaker of that particular language (see Acknowledgements). A complete list of the references used in collecting the dataset is provided in Online Appendix 1.

The data cover three standardized basic vocabularies found in the literature: the 100-word Swadesh list (Swadesh 1952), the 200-word Swadesh list (Swadesh 1955) and the Leipzig-Jakarta list (Tadmor 2009). The lists overlap to a considerable degree: the 100-word Swadesh list includes 93 meanings that are also in the 200-word Swadesh list, and the Leipzig-Jakarta list — which includes 100 meanings determined to represent basic vocabulary — shares 81 of its meanings with the 100-item Swadesh list and the 200-item Swadesh list, and consequently has nineteen meanings not found on either of the Swadesh lists. The total number of unique meanings in the full dataset is 226.

The presence or absence of a shared origin for the word representing each meaning in different languages is the main information in the dataset. This includes cases in which items are 'cognate' (i.e. vertically inherited from the same language of origin), as well as some cases where the items have been transmitted through borrowing. Thus, the presence-absence patterns do not strictly represent 'cognate' sets, but rather 'historical connections' (Kessler 2001) or 'correlates' (McMahon et al. 2005) that also include occasional borrowings tracing back to the same historical source. Because Uralic languages are well researched with respect to their historical connections, we were able to isolate the material further into qualitatively different kinds of lists, using information from the same reference sources that were used in compiling the dataset and determining the historical connections (Online Appendix 1). Each meaning in the dataset is accompanied by information showing any attested cases of borrowing, so that we could identify the meanings that to the best of our knowledge represented true cognate relationships and create a separate subset from them. As 100 of the 226 meanings matched this 'true cognate' criterion, we named this stable subset 'Ura100'. Despite the fact that Ura100 is of the same size as Swadesh100 and Leipzig-Jakarta, 34 of its 100 meanings are not on either of these lists.

Using the same loanword information used to isolate Ura100, we also isolated subsets with varying degrees of borrowing with which we could then investigate the effects of borrowings on the classification more tangibly than with standard basic vocabularies. For this we counted the number of attested borrowing events for each meaning outside Ura100 to arrive at an estimate of the borrowing-susceptibility of each meaning. For instance, the meaning “all” has two attested borrowing events: an apparent borrowing of an original Baltic (Indo-European) word into Finnic and a later borrowing of this once-borrowed word from a Finnic source language into a Saami language. Using this information we isolated five more sublists based on the number of attested borrowings: meanings with one or more borrowings (1+ borrowings, 124 items), meanings with two or more borrowings (2+ borrowings, 69 items), meanings with three or more borrowings (3+ borrowings, 47 items), meanings with four or more borrowings (4+ borrowings, 32 items) and meanings with five or more borrowings (5+ borrowings, 22 items). To maximize the difference between these lists and Ura100, only those meanings where the instances of borrowing were regarded as fairly certain were included. This is why, for instance, the number of meanings on the 1+ borrowings is 124 instead of 126, which would include all the meanings outside Ura100.

The basic vocabulary lists analysed are given in Table 3. For the analyses we used a 207-meaning Swadesh list (essentially a combination of the two Swadesh lists) instead of the 200-meaning Swadesh list. A more detailed description of the data is given in Online Appendix 2, which enumerates the meanings belonging to each subset.

Finally, in addition to present-day languages, we included reconstructed words representing Proto-Uralic, which we could use as the ‘outgroup’, a clade standing for an ancestor in the phylogenetic analyses and whose primary function

Table 3. An overview of the ten investigated sets of basic vocabulary meanings.

Set	Meanings	Includes
Full	226	All meanings in the dataset
Swadesh207	207	Meanings from the 200-word Swadesh list (Swadesh 1952) + 7 meanings from the revised Swadesh list (Swadesh 1955)
Swadesh100	100	Meanings from the revised Swadesh list (Swadesh 1955)
Leipzig-Jakarta	101	100 meanings with the highest rank from the Leipzig-Jakarta list of basic vocabulary (Tadmor 2009). The list we used included 101 meanings because the item ‘foot/leg’ in the original list was split into ‘foot’ and ‘leg’ as in Swadesh207.
Ura100	100	Meanings with no attested borrowings according to the references employed
1+ borrowings	124	Meanings with 1 or more borrowings
2+ borrowings	69	Meanings with 2 or more borrowings
3+ borrowings	47	Meanings with 3 or more borrowings
4+ borrowings	32	Meanings with 4 or more borrowings
5+ borrowings	22	Meanings with 5 or more borrowings

was to determine the position of the root for the phylogeny (Futuyma 1998: 97).[8] This Proto-Uralic was based on three primary sources: Itkonen & Kulonen (1992–2000), Rédei (1988–1991) and Sammallahti (1988). Significantly, the criteria for accepting items for this reconstruction differed from the criteria employed in mainstream Uralistics, where a reconstruction requires matches in any language of the two traditional 'main' branches, Finno-Ugric and Samoyed. This dichotomy is questioned by some Uralicists (see §2). To discourage the phylogenetic analysis from placing the root between Samoyed and Finno-Ugric only on the basis of the usual selection criteria of proto-forms, we included in our 'strict Proto-Uralic' only reconstructed forms that had likely counterparts in all three traditional deep-level subgroups: Ugric, Finno-Permian and Samoyed. This allowed for better alternative placements for the root, including between Finno-Permian and 'Ugric + Samoyed', one placement put forward in recent work (e.g. Häkkinen 2009).

5. Method

5.1 Bayesian phylogenetics

While various computational methods for reconstructing phylogenies have been developed, many recent studies rely on 'model-based' methods, which use a model specifying how a given change occurs in the data under investigation (in our case, the replacement of a correlate word with a non-correlate word or vice versa). Bayesian phylogenetic methods are among the most popular model-based methods today and are used here. A brief explanation of the method is given below; for further information in a less technical form, see Archibald et al. (2003) and for a more technical description, see Holder & Lewis (2003).

Bayesian methods calculate the statistical likelihood of the existence of certain kinds of results, such as the relationships of a set of languages. A substitution model defines how the data are likely to change in the course of time and the kinds of substitution possible. Given a substitution model and appropriately encoded language data, the Bayesian phylogenetic analysis produces a distribution of trees using a method called a Markov Chain Monte Carlo (MCMC) simulation. The simulation begins with a random tree shape and random values for the substitution model, for which the method calculates its likelihood (how well the values

8. The root could have been determined by other means, e.g. with software capable of rooting the phylogenies on independent grounds, such as BEAST. We refrained from trying out different rooting methods, which would certainly make an interesting topic in and of itself. Independent rooting has been recently applied to one basic vocabulary list in Honkola et al. (2013).

and the tree shape explain the data) and stores this state. This starting point will generally have a very poor likelihood associated with it. The algorithm then proposes new values and tree shapes in successive iterations; for each new value/tree shape combination proposed, the likelihood is recalculated. If the likelihood is better than that of the previous state, then the new state is accepted; if the likelihood of the new state is poorer than the previous iteration, then the new state is accepted in proportion to the drop in likelihood (i.e. the probability of the new state being accepted is high if the likelihood is only slightly lower, but the new state will probably not be accepted if its likelihood is very much lower than the previous state). The frequency with which the algorithm proposes new substitution model values or new tree shapes, as well as the actual values, can be influenced with 'priors', which essentially reflect what we claim to know of the system we are studying a priori and which can be very specific or very unspecific. Most Bayesian phylogenetic analyses use unspecific priors, as we do not want to influence the results with our own biases. The process of proposing new states is repeated (usually millions of times), and as the analysis progresses, each iteration converges on the most likely explanation(s) for the data. A portion of the trees produced in the process is sampled, usually at regular intervals, and the distribution of trees and values for the substitution model are the final results of the analysis.

5.2 Interpreting the results

The distribution of trees that the Bayesian phylogenetic analysis produces is summarized as 'majority-rule consensus trees', trees that include all the binary branchings occurring in more than 50% of all the trees in the distribution. Branchings that occur in less than 50% of the trees in the distribution are essentially random when it comes to an exact branching order and are thus represented as 'polytomies', branches splitting directly into more than two sub-branches. Majority-rule consensus trees contain two measures that are useful in interpreting our results. The first of these is a frequency of occurrence value called 'posterior probability', which ranges between 0.5 (50%) and 1 (100%); the other measure is branch length, which in 'phylograms', the type of trees shown here, approximates the number of substitutions per site over time and thus essentially reflects the amount of change along each branch.

Even if the majority-rule consensus tree provides a fair approximation of the most likely tree to explain the data, its branching events should be interpreted with care, with more confidence placed in the well-supported branchings (that is, branchings with a high posterior probability and a long branch length) than in the poorly supported branchings, which are ambiguous, sensitive to small changes in the data or both. As a general guideline often used for interpreting the kind of phylograms produced here (e.g. Huelsenbeck et al. 2001), values above 0.95 are considered to have very good support. Values slightly below this may also be noteworthy, but are less conclusive. In discussing the results we focus primarily on the branchings with support values ranging from 0.9 to 1.0 and maintain the general guideline that values of 0.95 or above are well supported, whereas those below this figure are tentatively supported. The weaker branchings may essentially be considered to be collapsed, meaning that the branching with the low posterior probability is removed and its constituent branches are attached to the well-supported branch below it, forming a 'polytomous' branch (a branch splitting into three or more sub-branches).

5.3 Details of the analyses conducted

The analyses were done with MrBayes, version 3.2.1 (Huelsenbeck & Ronquist 2001, Ronquist & Huelsenbeck 2003). Each analysis ran for 1 million generations, with every 1,000th generation sampled. The first 10% of trees were discarded as 'burn-in', to minimize any possible bias that the initial values might have on results. The substitution model we used for the analyses was a Markov K substitution model, a generalization of the JC69 (Jukes-Cantor) sequence evolution model. It assumes equal base frequencies and an equal probability for changes in either direction (Lewis 2001).

The analyses covered ten basic vocabulary lists: the full list of 226 meanings, Swadesh207, Swadesh100, Leipzig-Jakarta, Ura100 and five subsets with borrowing-susceptible items in different degrees (see §4). These meaning lists were converted into binary characters that correspond with etymological relationships, with the meanings belonging to a given correlate set marked as 1, meanings not belonging to a given correlate set marked as 0, and missing characters (i.e. the small number of meanings whose presence or absence in a language could not be ascertained from the references) marked with a question mark. An alternative solution would have been to use multi-state characters (for critical comparison between these alternatives, see Atkinson & Gray 2005: 104–105). The final results are presented as majority consensus trees.

6. Results

6.1 The full list of 226 meanings and the Swadesh207 list

The full list of meanings and the Swadesh207 list (Figure 5a, 5b) share seven clearly supported subgroups (see Tables 1–2): Samoyed, Permian, Ob-Ugric, Ugric,

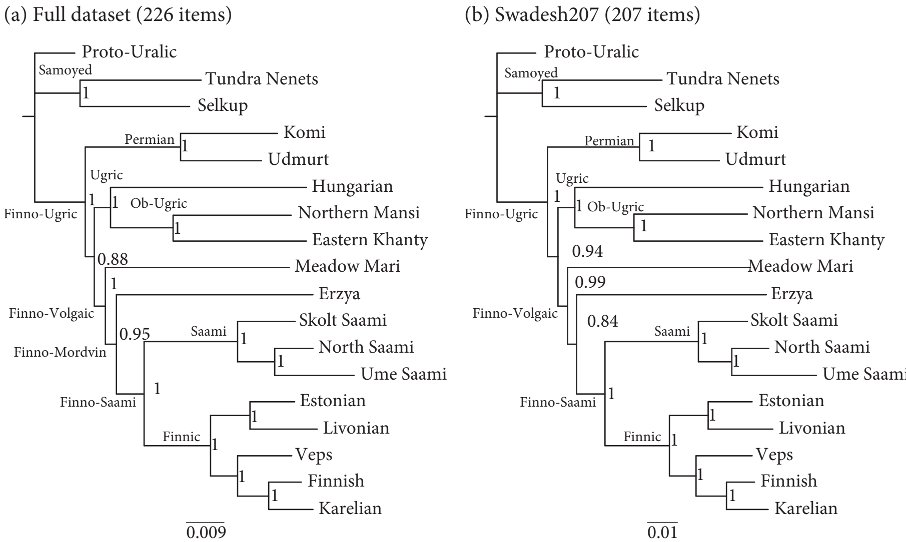


Figure 5. Phylograms produced from (a) the full dataset and (b) Swadesh207 subset. Posterior probability values in the nodes of the branches show strong (1–0.95), tentative (0.94–0.9) or low (< 0.9) support for the particular subgroup. Values below 0.5 are collapsed by the program and are not shown. Familiar intermediate subgroupings are labeled here but not in the remaining phylograms. The branch length reflects the amount of change.

Saami, Finnic and Finno-Saami. Further, Finnic is separated into a southern subgroup (Estonian + Livonian) and a northern one (Veps + Finnish + Karelian), and Saami into a western subgroup (North Saami + Ume Saami) and an eastern one (Skolt Saami). The Proto-Uralic outgroup, and consequently the initial split of these languages, is positioned between Samoyed and Finno-Ugric, which also holds for most of the other vocabulary lists.

The posterior probabilities of the phylograms are generally fairly high, with only one branch below 0.9 in each: Ugric + Finno-Volgaic in the full list, and Erzya + Finno-Saami (Finno-Mordvin) in Swadesh207. The branch lengths show that the branchings within Finno-Ugric that take place before Finno-Saami are supported by fewer characters than the other groupings generally, making them more sensitive to small variations in the data. To some extent this suggests that the diversification around this point may have been rapid, leaving limited traces in the basic vocabulary. In any case the short branch lengths along with the lower posterior probabilities suggest collapsing the phylogeny into three-way splits at two positions: Finno-Ugric, which would split into Permian, Ugric and Finno-Volgaic, and Finno-Volgaic, which would split into Meadow Mari, Erzya and Finno-Saami.

6.2 100-item basic vocabularies: Swadesh100 and Leipzig-Jakarta

The Swadesh100 and Leipzig-Jakarta lists (Figure 6a, 6b) classify Uralic languages similarly despite the fact that they share only 62 meanings. The only noteworthy difference in the branching order of the phylograms is in the northern Finnic subgrouping (Veps + Finnish + Karelian). Swadesh100, like Swadesh207 and the full list, suggests Veps as the first to diverge, while Leipzig-Jakarta suggests Finnish. Another difference between these two phylograms is the slightly lower support value given for the southern Finnic subgroup (Estonian + Livonian) in Swadesh100.

The basic subgroupings in the Swadesh100 and Leipzig-Jakarta phylograms are similar to those seen in the full list and Swadesh207. The main difference is the placement of Meadow Mari, which was positioned alongside Erzya and Finno-Saami in the larger lists, forming a Finno-Volgaic subgrouping. In contrast, Swadesh100 and Leipzig-Jakarta place Meadow Mari before Ugric and Finno-Mordvin, forming an unorthodox branch. However, the posterior probabilities of this branch and the subsequent Ugric + Finno-Mordvin branch are low. The branch lengths of the subgroupings that ensue from Finno-Ugric until around Finno-Mordvin are also short, suggesting a rapid or otherwise unclear diversification of the first branches within Finno-Ugric. Thus collapsing these branches is justified, giving a four-way split of Finno-Ugric into Permian, Meadow Mari, Ugric and Finno-Mordvin.

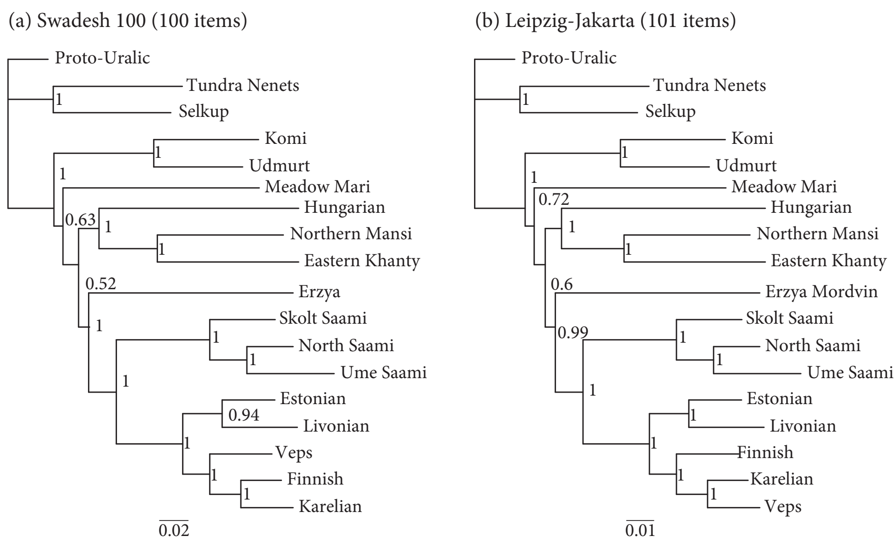


Figure 6. Phylograms produced from (a) Swadesh100 subset and (b) Leipzig-Jakarta subset. For further explanation, see Figure 5.

6.3 Uralic-specific basic vocabulary (Ura100)

The 100 items without documented borrowings (Ura100) appear to yield a unique classification with unusual subgroupings (Figure 7). Upon closer inspection, however, it becomes fairly similar to the four phylograms discussed above. Most of the common subgroupings seen in the other trees are again present with good support. The Ugric clade, however, is given a subpar support value. An unusual Permian + Ugric subgrouping is also present, although it has much too low a support value to be convincing. As with the previous phylograms, we can see a number of short intermediary branches, which along with the low posterior probabilities suggest collapsing the Finno-Ugric clade four ways into Permian, Hungarian, Ob-Ugric and Finno-Volgaic. It is also noteworthy that no definite branching order is suggested for Meadow Mari, Erzya or Finno-Saami.

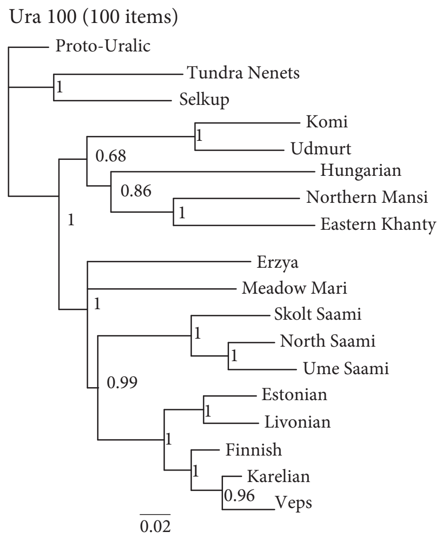
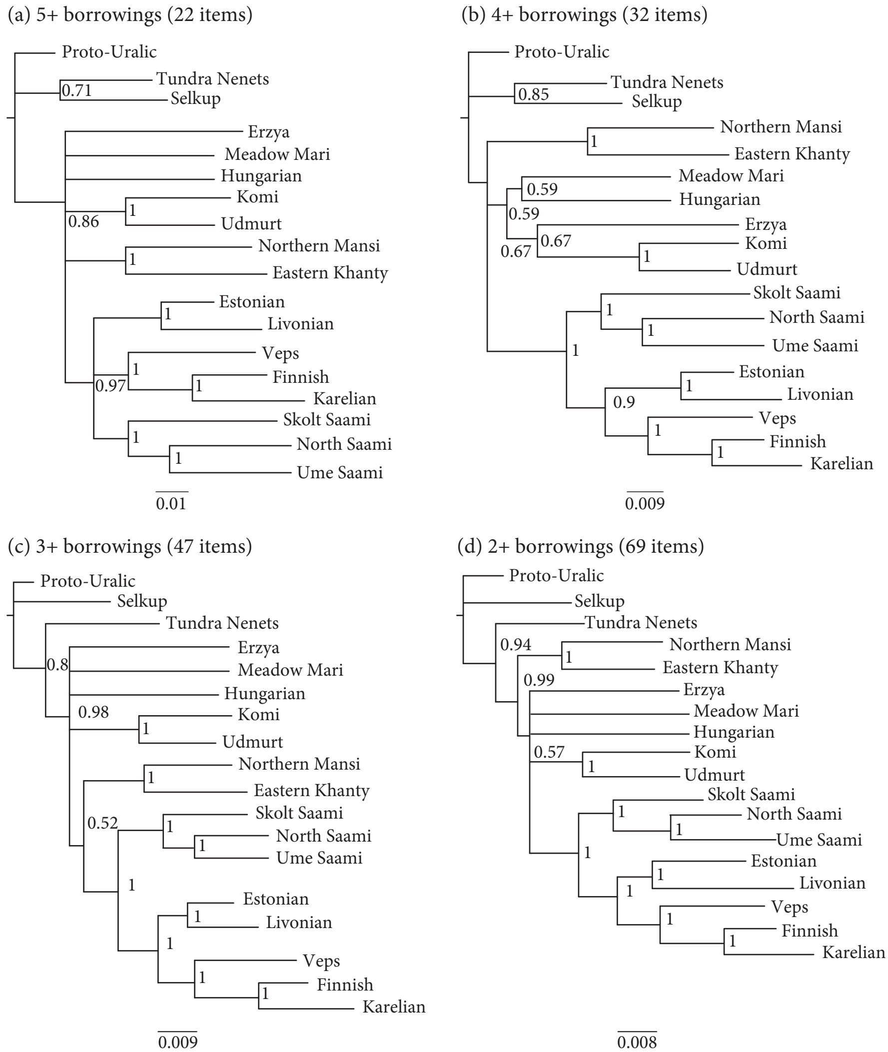


Figure 7. Phylogram produced from Ura100 subset, comprising only those meanings without attested borrowings in the languages under investigation. For further explanation, see Figure 5.

6.4 Meanings susceptible to borrowing

We conclude our discussion of the results by examining the five subsets all of whose meanings include some borrowing relationships (Figure 8a-e). The smallest of these subsets (5+ borrowings, Figure 8a) gives a polytomous, unresolved phylogeny. While borrowings may be a contributing factor, the small size of the subset (22 meanings) also affects the result, as the algorithm had little material to



work with (see §7.2). Collapsing the branches with posterior probabilities below 0.9 splits the phylogeny eight ways from the root. We can also identify a well-supported Saami and the northern and southern subgroups of Finnic that recur in the phylograms above. All are classified together as a Finno-Saami clade, albeit one without a clear separation of Saami and Finnic.

The slightly larger subset of 4+ borrowings, with 32 meanings (Figure 8b), also gives a polytomous classification, which (after accounting for weak support values) shows the same eight clades diverging from the root as the 5+ borrowings list. The groupings identified remain unchanged from 5+ borrowings with one

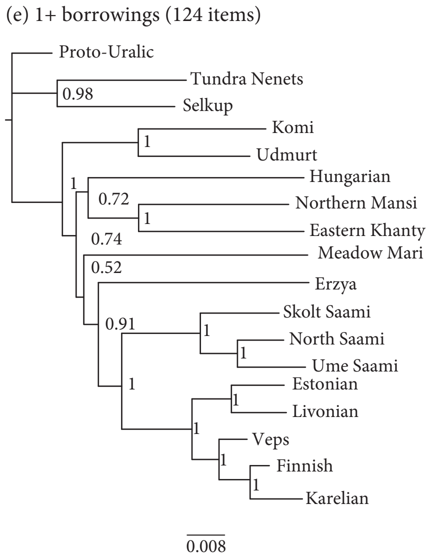


Figure 8. Phylogenies of borrowing-susceptible meanings in five partially overlapping subsets: (a) meanings with five or more borrowings (5+ borrowings), (b) meanings with four or more borrowings (4+ borrowings), meanings with three or more borrowings (3+ borrowings), (d) meanings with two or more borrowings (2+ borrowings) and (e) meanings with one or more borrowings (1+ borrowings). For further explanation, see Figure 5.

exception: the two subgroups of Finnic are now placed together, although only with tentative support.

3+ borrowings, with 47 meanings (Figure 8c), is the first subset to separate the Finno-Ugric languages from the two Samoyed languages (although the Samoyed languages are not grouped together). This reduces the initial division from eight clades to three if the weakly supported Tundra Nenets + Finno-Ugric branching is collapsed. Finno-Ugric remains fairly flat, dividing into six clades (Erzya, Meadow Mari, Hungarian, Permian, Ob-Ugric and Finno-Saami). In this phylogeny, the support for Finnic increases to 1.00.

The phylogram of the 69-meaning list of 2+ borrowings (Figure 8d) remains largely similar to the phylogram produced from the 3+ borrowings subset. Again, we end up with a three-way split into Finno-Ugric, Tundra Nenets and Selkup, although the order in the phylogram (Selkup splitting first, followed by Tundra Nenets and then Finno-Ugric) has only tentative support. This phylogeny yields a flat Finno-Ugric divided into the same six clades as 3+ borrowings.

Superficially, the largest subset with 124 meanings (1+ borrowings, Figure 8e) gives a fairly good match with either the full list phylogram or the Swadesh207 phylogram. However, after taking the posterior probabilities into account, we end up with a five-way division of Finno-Ugric (into Permian, Hungarian, Ob-Ugric, Meadow Mari and Finno-Mordvin). The Samoyed languages are also grouped together with a very good support value. In addition, the first immediate subgroups of Finno-Ugric have the shortest branch lengths, which again resemble the classifications produced by the standardized basic vocabulary lists.

7. Discussion

7.1 Basic vocabulary classification of the Uralic languages

As established in §5.2, a good rule of thumb in interpreting the support values is to regard those exceeding 0.95 as well supported (Huelsenbeck et al. 2001); those with slightly smaller support values may also be worth examining, but they are more tentative. With respect to these guidelines, Table 4 summarizes the subgroupings that occur in the phylogenies with a support value of 0.95 ± 0.05. The subgroupings are placed in descending order according to the sum of their

Table 4. Subgroupings from the phylograms with support values (posterior probabilities) of 0.95 ± 0.05.

	Full list (226)	Sw207 (207)	Sw100 (100)	LJ (101)	Ura100 (100)	5+ B. (22)	4+ B. (32)	3+ B. (47)	2+ B. (69)	1+ B. (124)
Ob-Ugric	1	1	1	1	1	1	1	1	1	1
Northern Finnic	1	1	1	1	1	1	1	1	1	1
Permian	1	1	1	1	1	1	1	1	1	1
Saami	1	1	1	1	1	1	1	1	1	1
Finno-Saami	1	1	1	1	0.99	0.97	1	1	1	1
Southern Finnic	1	1	0.94	1	1	1	1	1	1	1
Finnic	1	1	1	1	1	–	0.9	1	1	1
Finnish + Karelian	1	1	1	–	–	1	1	1	1	1
Finno-Ugric	1	1	1	1	1	–	–	0.98	0.99	1
Samoyed	1	1	1	1	1	–	–	–	–	0.98
Ugric	1	1	1	1	–	–	–	–	–	–
Finno-Mordvin	0.95	–	1	0.99	–	–	–	–	–	0.91
Finno-Volgaic	1	0.99	–	–	1	–	–	–	–	–
Karelian + Veps	–	–	–	1	0.96	–	–	–	–	–
Ugric + Finno-Volgaic	–	0.94	–	–	–	–	–	–	–	–
Tundra Nenets + Finno-Ugric	–	–	–	–	–	–	–	–	0.94	–

posterior probabilities across all ten phylograms. Thus, the table provides a rough approximation of the overall level of confidence for each branching with at least tentative support. This should also make it easier to compare the results of the analyses. Note, however, that the order of elements in Table 4 does not account for the branch lengths.

It is generally argued that lexical evidence in particular supports the traditional root position between Finno-Ugric and Samoyed (e.g. Salminen 2002, cf. Janhunen 2000). Our results were in line with this suggestion, as this root position was the most frequent, occurring in six of the ten phylograms, but it was absent from 5+ borrowings, 4+ borrowings, 3+ borrowings and 2+ borrowings. The most interesting deviation from the orthodox root position is seen in the 2+ borrowings list, where the root was positioned with tentative support between Selkup and Tundra Nenets + Finno-Ugric. Finno-Ugric disappears in only the two smallest and least 'optimal' sublists, namely, 4+ borrowings and 5+ borrowings. Still, if we also look at the phylograms that do not reconstruct Samoyed or Finno-Ugric, we find that even there the root position remains very close to the textbook dichotomy between Finno-Ugric and Samoyed, although the 'strict Proto-Uralic' outgroup should not encourage this root position (see §4).

Moving one step towards the terminal branches, we encounter the feature that (perhaps) stands out most in comparison with orthodox Uralic classifications — the position of Ugric. In our results Ugric is regularly placed either alongside Permian or beside Finno-Volgaic and after Permian, but never before it. Consequently, all the phylograms lack the traditionally attested Finno-Permian subgrouping that would follow the separation of Ugric from Finno-Ugric (e.g. Figure 2), with the Swadesh207 phylogram even giving tentative support to reversing the traditional positions of Permian and Ugric, creating an Ugric + Finno-Volgaic branch. While this unorthodox subgrouping remains unconvincing, with only Swadesh207 suggesting it, its presence along with the short branch lengths across all the phylograms argues against a discrete Finno-Permian stage,[9] instead implying a polytomous divergence of Finno-Ugric. The results give slightly more support for maintaining two often-cited levels that follow Finno-Permian: Finno-Volgaic and Finno-Mordvin. Finno-Volgaic is well supported in three phylograms (the full list, Swadesh207 and Ura100), Finno-Mordvin is present in four phylograms (the full list, Swadesh100,

9. However, the short branch lengths and the low posterior probabilities around this point also mean that the branching order could easily change in favour of Finno-Permian if the analyses were repeated using different data. While the weakness of the Finno-Permian branch has been pointed out by Salminen (2002) on the basis of scarce phonological innovations, in other sources Finno-Permian is considered to be well supported in other parts of the language, including specialized vocabulary and grammar (e.g. Décsy 1965: 181–183).

Leipzig-Jakarta and 1+ borrowings). However, these two levels are also partially unclear, as their branch lengths remain fairly short and are thus supported by few lexical innovations in the basic vocabulary. To sum up, most of the ambiguity is found around the first clades of Finno-Ugric, particularly in the order of Ugric and Permian, but also to some extent around Finno-Volgaic and Finno-Mordvin. These positions are among the hardest to place on a bifurcating tree.

As Table 3 shows, the Ob-Ugric subgrouping, consisting of Mansi and Khanty, has full support in all phylograms. The Ugric branch, consisting of Hungarian and the Ob-Ugric languages, is present with good support in four phylograms: the full list, Swadesh207, Swadesh100 and Leipzig-Jakarta; with the loan-free Ura100 list, the support value falls below 0.9 (Figure 7). Based on the results, Ugric has surprisingly weak support from basic vocabulary, roughly equal to that of Finno-Mordvin.[10]

The Volgaic group (Mari + Mordvin) is completely absent from the phylograms. Meadow Mari and Erzya are generally placed on the same branch or on adjacent branches so that Erzya is one branch closer to Finno-Saami than Meadow Mari. The results thus agree with the current view that the linguistic similarities between these languages reflect primarily areal connections, and that the genealogical connection between them runs only as deep as it does between them and the other Finno-Volgaic languages (Michalove 2002). Further, the areal connections between the two languages are not intensive enough to emerge in any of the loanword subsets, suggesting essentially separate linguistic histories for these two languages after their divergence from the Finno-Volgaic stage.

Finno-Saami, which is generally accepted today, is also well supported in all the basic vocabulary phylograms (Table 3), and it also has a long branch length. The hierarchy separating Saami from Finnic is also highly regular; regardless of whether we look at stable basic vocabulary or borrowing-susceptible items, the hierarchy of Finnic breaks down only in the smallest and most borrowing-susceptible subset (5+ borrowings), which still shows the Finno-Saami subgrouping, but now with a three-way split into Saami, Southern Finnic and Northern Finnic (Table 3, Figure 8a). This sublist is more likely to reflect the fairly strong flow of borrowings between Finnic and Saami, which could explain this result. However, as the 5+ borrowings list also consists of a very small number of meanings, which in itself reduces the accuracy of the classification, the size in addition to the conflicting historical patterns may be the reason for the change in the internal hierarchy. Thus, our results give high reliability to Finno-Saami.

10. However, it should be noted that numerous characteristics outside the material explored here have been taken to firmly support Ugric, including common grammatical innovations and lexicon outside the basic vocabulary (Honti 1998).

The results show that the basic vocabulary regularly separates the Finnic languages into a southern group, consisting of Estonian and Livonian, and a northern group, consisting of Finnish, Veps and Karelian. This kind of division has been demonstrated in the literature (Turunen 1988), but is not the only one suggested; Laakso (2001: 204–207) notes isoglosses that suggest not only a division on a north-south axis, but also on an east-west axis. While placing these languages on a tree diagram should thus not be very straightforward, the basic vocabulary analyses show surprisingly little variation. However, while the north-south division appeared to be fairly straightforward, ambiguity on a smaller scale is apparent, particularly when comparing the northern Finnic subgrouping (Finnish + Veps + Karelian) in the different phylograms. These show two patterns: a more frequent one, found in all phylograms except Leipzig-Jakarta and Ura100, where Veps is the first language to separate, and a less frequent one, found in Ura100 and Leipzig-Jakarta, where Finnish is first to separate. This pattern may be attributable to the geographical dialect continuum across Finnish, Karelian and Veps. Thus, the division of Finnic languages into southern and northern groups is firm enough to be indisputable in the phylogenetic analyses, while the internal division of these groups becomes too shallow, yielding contradictory results.

The Saami subgrouping is even more consistent than Finnic (Table 3). The languages were divided in all phylograms into Skolt Saami and North Saami + Ume Saami, thus matching a division into eastern and western languages, similar to e.g. Korhonen (1981: 18). This division has usually been based mainly on phonological and morphological features (Sammallahti 1998: 6–7), but as seen here, appears replicable also with basic vocabulary data. Interestingly, while the Saami languages have a considerable amount of attested borrowing activity, particularly with Finnic, the branching order remains the same in all the phylograms.

In summary, the basic vocabulary analyses suggest that the root of the Uralic phylogeny belongs quite firmly between Samoyed and Finno-Ugric. The Finno-Ugric level is followed by the greatest degree of ambiguity, reflected in our results by weak posterior probabilities, short branch lengths and variation in the branching order of the different subsets. This would be the most plausible place to collapse the phylogeny should one desire to do so. The results show that the basic vocabulary does not support the Finno-Permian clade and gives only slightly higher support for the subsequent Finno-Volgaic and Finno-Mordvin clades. The traditional Ugric clade is also shown as somewhat weak, but nonetheless partially supported. From Finno-Saami onwards, the results again become very regular, giving us a firm Finno-Saami stage, a north-south division for Finnic languages and an east-west division of the Saami languages. More or less all of the phylograms with 100 or more items collapse into the kind of classification described above after the posterior probabilities are taken into account.

7.2 Qualitative and quantitative differences in the word lists / phylograms

The subgroupings summarized in Table 3 show that the number of meanings used in the analysis roughly correlates with the number of well-supported subgroupings that can be identified from the phylograms: 200 or more meanings (the full list and Swadesh207) produce thirteen subgroupings, 100 to 124 meanings (Swadesh100, Leipzig-Jakarta, Ura100, 1+ borrowings) produce eleven or twelve subgroupings (Swadesh100 and Leipzig-Jakarta have one subgrouping more than Ura100 and 1+ borrowings), 69 meanings (2+ borrowings) produce ten subgroupings, 47 meanings (3+ borrowings) produce nine subgroupings, 32 meanings (4+ borrowings) produce eight subgroupings, 22 meanings (5+ borrowings) produce seven subgroupings.[11] Thus, when it comes to quantitative differences, the number of items analysed more or less directly reflects how 'resolved' the resulting phylogeny is (that is, how many well-supported bifurcations it contains), although there is some fluctuation as the lists with 100 to 124 meanings show. The overall result in any case essentially reflects Embleton's (1986) point that more items generally produce more accurate results.

When it comes to qualitative differences, in particular the effect of borrowings, changing from one basic vocabulary list to another does not appear to have a significant effect on how the Uralic languages are positioned or (for the most part) what subgroupings are identified, although a fairly large proportion (55.7% of the 226 encoded meanings) contain at least some borrowing activity, which is also reflected in their encoding. The subgroupings most prone to change include several that have been deemed problematic, including Ugric, Finno-Mordvin and Finno-Volgaic, which fluctuated between Ura100 and the basic vocabularies. Interestingly, the 1+ borrowings list, which we may also roughly place in the 100-item category, produces less certain results (low posterior probabilities), but ones that are nonetheless quite similar to Swadesh100 or Leipzig-Jakarta, except that a well-supported Ugric clade is not among them. Considering that the substitution model employed is fairly simple, the most likely interpretation for this similarity

11. As it was uncertain how dramatic an effect the qualitative differences (the number of borrowings) in the five borrowing-susceptible lists had, we conducted an additional experiment by randomly selecting Ura100 meanings, producing subsets with 22 meanings, 32 meanings, 47 meanings and 69 meanings — reflecting the sizes of 5+, 4+, 3+ and 2+ borrowings, respectively — analyzed them with MrBayes and counted the number of subgroupings with tentative or good support (data not shown). These yielded the same number of subgroupings as the borrowing lists of the same size, with the exception that 69 meanings from Ura100 produced a phylogeny with the same subgroupings as Ura100 in its entirety, unlike the 2+ borrowings list, which remained less resolved, as is apparent from Table 3. These analyses support the outcome that the size of the dataset is crucial for this method to produce a resolved tree.

would not be that the method is so efficient in working around conflicts in the material, but rather that the pattern(s) emerging from the borrowings within the Uralic basic vocabulary do not conflict on the whole with the patterns emerging from the non-borrowed basic vocabulary items. The matter of borrowings and their effects on the Uralic phylogeny could be examined more extensively with less basic vocabulary than used here.

In general, the number of loans does affect the phylogram to an extent (e.g. the Ura100 list vs. 1+ borrowings), but on the other hand, the larger datasets appear to produce quite similar results regardless of the presence of borrowings (Ura100 vs. the full list). Quantity has a significant role in determining the shape of the classification; with a smaller number of items the phylogram becomes flatter and less robust, and with lists such as 4+ borrowings and 5+ borrowings, it is essentially without hierarchy.

8. Conclusion

From the perspective of Bayesian phylogenetic methods and quantitatively and qualitatively different basic vocabulary lists, the Uralic language family appears to remain more regular than the considerable discussion in Uralistics and the alternative classifications from recent decades would lead one to believe. Our results agree more with a traditional classification along the lines of Korhonen (1981) than with the more polytomous classification models. However, both the posterior probabilities and the branch lengths point to ambiguity in the first branchings of Finno-Ugric, which would justify collapsing some branches around this point. Particularly the absence of the Finno-Permian clade differentiates our results from many previous classifications. Notably, the results are very similar in terms of classification as those in Honkola et al. (2013), which examined the Ura100 dataset with the BEAST software package and a different model.

The number of items is roughly reflected in the number of well-attested subgroupings, and larger datasets continuously lead towards more resolved phylogenies. However, the classification changes surprisingly little with loanword meanings: the phylogram produced from the borrowing-susceptible basic vocabulary items is very similar to the results of the standardized basic vocabulary lists or to a list of non-borrowed basic vocabulary. Consequently, borrowings within the Uralic basic vocabulary appear to have rather little effect on the classification, implying either that the borrowings among the basic vocabulary items are only a marginal hindrance for the contemporary methods in general or, what is perhaps more likely, that the borrowings within the Uralic basic vocabulary do not substantially conflict with the less borrowing-susceptible items. The agreement between the more borrowing-susceptible and the less borrowing-susceptible datasets would be expected if the borrowings occurred mainly between close relatives, which would appear to be the case with the geographically widespread Uralic languages.

References

Abondolo, Daniel (ed.). 1998. *The Uralic languages.* London: Routledge.

Alpher, Barry & David Nash. 1999. Lexical replacement and cognate equilibrium in Australia. *Australian Journal of Linguistics* 19. 5–56.

Archibald, Jenny K., Mark E. Mort & Daniel J. Crawford. 2003. Bayesian inference of phylogeny: A non-technical primer. *Taxon* 52. 187–191.

Atkinson, Quentin D. & Russell Gray. 2005. Curious parallels, curious connections — phylogenetic thinking in biology and historical linguistics. *Systematic Biology* 54(4). 513–526.

Atkinson, Quentin D., Geoff Nicholls, David Welch & Russell Gray. 2005. From words to dates: Water into wine, mathemagic or phylogenetic inference? *Transactions of the Philological Society* 103(2). 193–219.

Austin, Peter K. (ed.). 2008. *1000 languages: The worldwide history of living and lost tongues.* London: Thames & Hudson.

Bastin, Yvonne, Andre Coupez & Michael Mann. 1999. *Continuity and divergence in the Bantu languages: Perspectives from a lexicostatistic study*. Tervuren, Belgium: Musée royal de l'Afrique centrale.

Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.

Bowern, Claire. 2010. Historical linguistics in Australia: Trees, networks and their implications. *Philosophical Transactions of the Royal Society B* 365. 3845–3854.

Croft, William. 2000. *Explaining language change: An evolutionary approach.* Essex: Pearson Education Limited.

Croft, William. 2006. The relevance of an evolutionary model to historical linguistics. In Ole Nedergaard Thomsen (ed.), *Competing models of linguistic change: Evolution and beyond*, 91–132. Amsterdam: John Benjamins.

Croft, William. 2008. Evolutionary linguistics. *Annual Review of Anthropology* 37. 219–234.

Crystal, David. 1987. *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.

Décsy, Gyula. 1965. *Einführung in die finnisch-ugrische Sprachwissenschaft*. Wiesbaden, Germany: Otto Harrassowitz.

Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesink & Angela Terrill. 2008. Structural phylogeny in historical linguistics: Methodological explorations applied in Island Melanesia. *Language* 84(4). 710–759.

Dyen, Isidore, A.T. James & J.W. Cole. 1967. Language divergence and estimated word retention rate. *Language* 43. 150–171.

Embleton, Sheila. 1986. *Statistics in historical linguistics.* Bochum, Germany: Brockmeyer.

Futuyma, Douglas. 1998. *Evolutionary biology*, 3rd edn. Sunderland, MA: Sinauer Associates.

Gray, Russell D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426. 435–439.

Gray, Russell D. & Fiona M. Jordan. 2000. Language-trees support the express-train sequence of Austronesian expansion. *Nature* 405. 1052–1055.

Greenhill, Simon & Russell D. Gray. 2009. Australian language phylogenies: Myths and misconceptions about Bayesian computational methods. In Alexander Adelaar & Andrew Pawley (eds.), *Austronesian historical linguistics and cultural history: A festschrift for Robert Blust*, 375–397. Canberra: Pacific Linguistics.

Greenhill, Simon, Thomas E. Currie & Russell D. Gray. 2009. Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society B*. 276. 2299–2306.

Häkkinen, Jaakko. 2009. Kantauralin ajoitus ja paikannus: perustelut puntarissa. *Journal de la Société Finno-Ougrienne* 92. 9–56.

Häkkinen, Kaisa. 1983. *Suomen kielen vanhimmasta sanastosta ja sen tutkimisesta*. Turku, Finland: University of Turku dissertation.

Häkkinen, Kaisa. 1984. Wäre es schon Zeit, den Stammbaum zu fällen? Theorien über die gegenseitigen Verwandtschaftsbeziehungen der finnisch-ugrischen Sprachen. *Ural-Altaische Jahrbücher, Neue Folge* 4. 1–24.

Holder, Mark & Paul O. Lewis. 2003. Phylogeny estimation: Traditional and Bayesian approaches. *Nature Reviews Genetics* 4. 275–284.

Honkola, Terhi, Outi Vesakoski, Kalle Korhonen, Jyri Lehtinen, Kaj Syrjänen & Niklas Wahlberg. 2013. Cultural and climatic changes shape the evolutionary history of the Uralic languages. *Journal of Evolutionary Biology*, 26. 1244–1253. http://onlinelibrary.wiley.com/doi/10.1111/jeb.12107/pdf. (5 June, 2013).

Honti, László. 1998. Ugrilainen kantakieli — erheellinen vai realistinen hypoteesi. In Riho Grünthal & Johanna Laakso (ed.), *Oekeeta asijoo. Commentationes Fenno-Ugricae in honorem Seppo Suhonen sexagenarii 16. V. 1998.* (Suomalais-Ugrilaisen Seuran Toimituksia 228), 176–187. Helsinki: Suomalais-Ugrilainen Seura.

Hovdhaugen, Even, Fred Karlsson, Carol Henriksen & Bengt Sigurd. 2000. *The history of linguistics in the Nordic Countries.* Helsinki: Societas Scientiarum Fennica.

Huelsenbeck, John P. & Fredrik Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17. 754–755.

Huelsenbeck, John P., Fredrik Ronquist, Rasmus Nielsen & Jonathan P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294. 2310–2314.

Itkonen, Erkki & Ulla-Maija Kulonen (eds.). 1992–2000. *Suomen sanojen alkuperä*, 1–3. Helsinki: Suomalaisen Kirjallisuuden Seura.

Itkonen, Terho. 1997. Reflections on Pre-Uralic and the 'Saami-Finnic protolanguage'. *Finnisch-Ugrische Forschungen* 54(3). 229–266.

Janhunen, Juha. 2000. Reconstructing Pre-Proto-Uralic: Spanning the millennia of linguistic evolution. In Anu Nurk, Triinu Palo & Tõnu Seilenthal (ed.), *Congressus Nonus Internationalis Fenno-Ugristarum*, vol. 1, 59–76. Tartu, Estonia: Eesti Fennougristide Komitee.

Janhunen, Juha. 2009. Proto-Uralic — what, where, and when? *Mémoires de la Société Finno-Ougrienne* 258. 57–78.

Kessler, Brett. 2001. *The significance of word lists.* Stanford: CSLI.

Korhonen, Mikko. 1981. *Johdatus lapin kielen historiaan.* Helsinki: Suomalaisen Kirjallisuuden Seura.

Kulonen, Ulla-Maija. 2002. Kielitiede ja Suomen väestön juuret. In Riho Grünthal (ed.), *Ennen muinoin: miten menneisyyttämme tutkitaan*, 102–116. Helsinki: Suomalaisen Kirjallisuuden Seura.

Künnap, Ago & Rein Taagepera. 2005. Distances among Uralic and other Northern Eurasian languages. *Linguistica Uralica* 3. 161–181.

Laakso, Johanna. 2001. The Finnic languages. In Östen Dahl (ed.), *The Circum-Baltic languages*, vol. 1: *Past and present*, 179–212. Amsterdam: John Benjamins.

Lee, Sean & Toshikazu Hasegawa. 2011. Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. *Proceedings of the Royal Society B.* 278. 3662–3669.

Lewis, Paul O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50(6). 913–925.

McMahon, April, Paul Heggarty, Robert McMahon & Natalia Slaska. 2005. Swadesh sublists and the benefits of borrowing: An Andean case study. *Transactions of the Philological Society* 103(2). 147–170.

McMahon, April & Robert McMahon. 2003. Finding families: Quantitative methods in language classification. *Transactions of the Philological Society* 101(1). 7–55.

McMahon, April & Robert McMahon. 2005. *Language classification by numbers*. Oxford: Oxford University Press.

Matisoff, James. 2000. On the uselessness of glottochronology for the subgrouping of Tibeto-Burman. In Colin Renfrew, April McMahon & Larry Trask (eds.), *Time depth in historical linguistics*, 333–371. Cambridge: The McDonald Institute for Archaeological Research.

Matthews, Stephen & Maria Polinsky. 1997. Europe and Eurasia. In Bernard Comrie, Stephen Matthews & Maria Polinsky (eds.), *The atlas of languages: The origin and development of languages throughout the world*, 36–55. London: Bloomsbury.

Michalove, Peter A. 2002. The classification of the Uralic languages: Lexical evidence from Finno-Ugric. *Finnisch-Ugrische Forschungen* 57. 58–67.

Nakhleh, Luay, Tandy Warnow, Don Ringe & Steven N. Evans. 2005. A comparison of phylogenetic reconstruction methods on an Indo-European dataset. *Transactions of the Philological Society* 103(2). 171–192.

O'Grady, Geoffrey N. 1960. More on lexicostatistics. *Current Archaeology* 1. 338–339.

Pagel, Mark. 2009. Human language as a culturally transmitted replicator. *Nature Reviews Genetics* 10. 405–415.

Pagel, Mark, Quentin Atkinson & Andrew Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449. 717–720.

Price, Glanville (ed.). 1998. *Encyclopedia of the languages of Europe*. Oxford: Blackwell.

Raun, Alo. 1956. Über die sogennante lexicostatistiche Methode oder Glottochronologie und ihre Anwendung auf das Finnisch-Ugrische und Türkische. *Ural-Altaische Jahrbücher* 28. 151–154.

Rédei, Károly. 1988–1991. *Uralisches Etymologisches Wörterbuch*, vols.1–3. Budapest: Akadémiai Kiadó.

Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100(1). 59–129.

Ronquist, Fredrik & John P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19. 1572–1574.

Ruhlen, Merritt. 1987. *A guide to the world's languages*, vol. 1: *Classification*. Stanford, CA: Stanford University Press.

Saarikivi, Janne. 2011. Saamelaiskielet — nykypäivää ja historiaa. In Irja Seutujärvi-Kari, Petri Halinen & Risto Pulkkinen (eds.), *Saamelaistutkimus tänään*, 77–119. Helsinki: Suomalaisen Kirjallisuuden Seura.

Salminen, Tapani. 1999. Euroopan kielet muinoin ja nykyisin. In Paul Fogelberg (ed.), *Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan*, 13–26. Helsinki: Societas Scientiarum Fennica.

Salminen, Tapani. 2002. Problems in the taxonomy of the Uralic languages in the light of modern comparative studies. In Александр Е. Кибрик (ed.), *Лингвистический беспредел: сборник статей к 70-летию А. И. Кузнецовой*, 44–55. Moscow: Издательство Московского университета.

Salminen, Tapani. 2007. Europe and North Asia. In Christopher Moseley (ed.), *Encyclopedia of the world's endangered languages*, 211–282. New York: Routledge.

Sammallahti, Pekka. 1988. Historical phonology of the Uralic languages with special reference to Samoyed, Ugric and Permic. In Denis Sinor (ed.), *The Uralic languages: Description, history and foreign influences*, 478–554. Leiden: Brill.

Sammallahti, Pekka. 1998. *The Saami languages: An introduction.* Kárášjohka, Norway: Davvi Girji.

Sankoff, David. 1973. Mathematical developments in lexicostatistical theory. In Thomas Sebeok (ed.), *Current trends in linguistics 11: Diachronic, areal and typological linguistics*, 93–112. The Hague: Mouton.

Sinor, Denis. 1988. Introduction. In Denis Sinor (ed.), *The Uralic languages: Description, history and foreign influences*, xiii–xx. Leiden: Brill.

Swadesh, Morris. 1952. Lexicostatistic dating of prehistoric ethnic contacts. *Proceedings of the American Philological Society* 96. 452–463.

Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21. 121–137.

Taagepera, Rein. 1994. The linguistic distances between Uralic languages. *Linguistica Uralica* 30. 161–167.

Tadmor, Uri. 2009. Loanwords in the world's languages: Findings and results. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages: A comparative handbook*, 55–75. Berlin: Walter de Gruyter.

Tambovtsev, Yuri. 2004. Uralic language taxon: Natural or artificial? (Typological compactness of Uralic languages taxons: Branches, subgroups, groups, families and superfamilies). *Fenno-Ugristica* 26. 200–246.

Turunen, Aimo. 1988. The Balto-Finnic languages. In Denis Sinor (ed.), *The Uralic languages: Description, method and foreign influences*, 58–83. Leiden: Brill.

Résumé

Nous avons étudié la diversification des langues finno-ougriennes en utilisant les méthodes numériques bayésiennes, qui sont bien établies dans l'analyse phylogénétique. Nous avons analysé dix ensembles différents issus du vocabulaire de base pour pouvoir évaluer si les sous-groupes de cette famille de langues se justifient. Selon nos résultats, la phylogénie ouralienne dérivée du vocabulaire de base a une topologie relativement solide qui subit peu d'altérations quand on utilise des séries de mots plus susceptibles d'être empruntés. Certaines différences existent entre nos résultats et le modèle le plus largement admis : par exemple, l'absence de preuves solides de l'existence d'un sous-groupe finno-permien.

Zusammenfassung

Motiviert von der andauernden Diskussion bezüglich der Klassifikation von uralischen Sprachen untersuchen wir diese Sprachfamilie quantitativ durch Anwendung der Bayesischen Methode der phylogenetischen Analyse. Wir untersuchen den Grundwortschatz von siebzehn Sprachen. Um die Robustheit der Sprachgruppen innerhalb dieser Sprachfamilie auszuwerten, analysieren wir zehn verschiedene Sets von Wörtern aus dem Grundwortschatz. Diese Sets beinhalten Wortlisten aus der Literatur, Wortlisten deren Wörter nicht durch Entlehnungen erkennbar sind, Wortlisten mit Wörtern die unterschiedlich leicht durch Entlehnung erkennbar sind, und eine Wortliste die all diese untersuchten Wortlisten enthalten. Unsere Ergebnisse zeigen, dass die uralische Phylogenie — aus der Grundwortschatzperspektive — eine stabile Gestalt annimmt, die von Entlehnungsbedeutungen nicht dramatisch verändert wird. Unsere Ergebnisse weichen teilweise von dem ‚Standardparadigma' Klassifikationsansatz dieser Sprachen ab, wie beispielsweise der Mangel für konkrete Beweise für die finno-permische Sprachgruppe.

Corresponding author's address

Kaj Syrjänen
School of Language, Translation and Literary Studies
FIN-33014 University of Tampere
Finland

kaj.jaakko.syrjanen@uta.fi