

RECONSTRUCTING THE LATERAL COMPONENT OF LANGUAGE HISTORY AND GENOME EVOLUTION USING NETWORK APPROACHES

*Shijulal Nelson-Sathi, Ovidiu Popa, Johann-Mattis List, Hans Geisler,
William F. Martin and Tal Dagan*

PREFACE

Genome evolution and the history of language development share many features. Both processes involve basic elements – words or genes – whose properties can change over time. An alteration of an element’s property can lead to a change in its function that in turn may affect the structure and composition of the whole domain, be it a language or a genome. Similarly to genomes that owe their existence to their corresponding species, languages also exist as long as there exists a population of native-speakers. Both genomes and languages may vary within the population. Eventually, the population that carries the domains – speakers or organisms – may split and continue to change independently, resulting in a divergence event and the origin of new languages or species.

First investigations into the similarities between language and species evolution are documented in the early modern period. These were the times when “Catastrophism”, the leading paradigm of natural history, linguistics, and geology during the Middle Ages and the Early Modern Period, lost its power (Wells 1973). Under the slogan, “The present is the key to the past,” which was originally coined by geologists (Cannon 1960), genealogical relations were inferred for species and languages. Shared traits in their present form were interpreted as evidence for their past identity, and the family tree became the leading metaphor to describe genealogical relations in linguistics as well as in biology.

As Geisler & List (this volume) point out, methods for phylogenetic reconstruction were developed independently in biology and linguistics, with August Schleicher (1821–1861) being among the first linguists to model language evolution by means of bifurcating trees (Schleicher 1853a, Schleicher 1853b, see Figure 1), and Charles Darwin (1809–1882) being among the first biologists to illustrate the splitting of ancestor species into their descendants with help of the family tree schema (Darwin 1859, see Figure 2). Neither Darwin nor Schleicher were the first to use trees to depict species or language evolution, yet both made the tree model popular in their respective disciplines (Ragan 2009; Sutrop 1999).

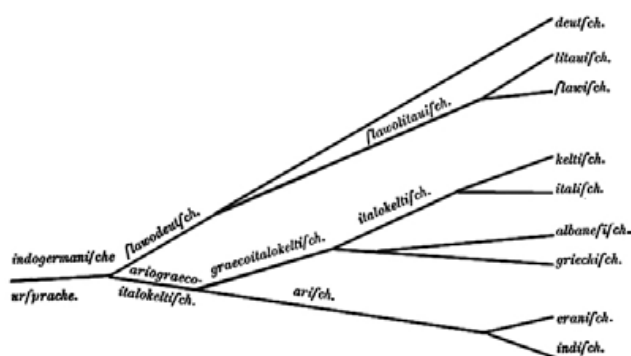


Figure 1: Family tree of the Indo-European languages by August Schleicher (1861:7)

Biologists and linguists soon became aware of some striking similarities, not only between the objects, but also between the processes that they were investigating. Darwin briefly addressed the topic of language evolution in his work (Darwin 1859), and August Schleicher devoted an essay to the German biologist Ernst Haeckel (1834–1919). The essay, dealing with parallels and differences between language classification and species evolution, was published as an open letter (Schleicher 1863). In his essay, Schleicher addressed explicitly the importance of the uniformitarian principle (ibid. 10f) and the family tree model (ibid. 14f) in both disciplines, and emphasized the differences between the biological and linguistic entities (Schleicher 1863).

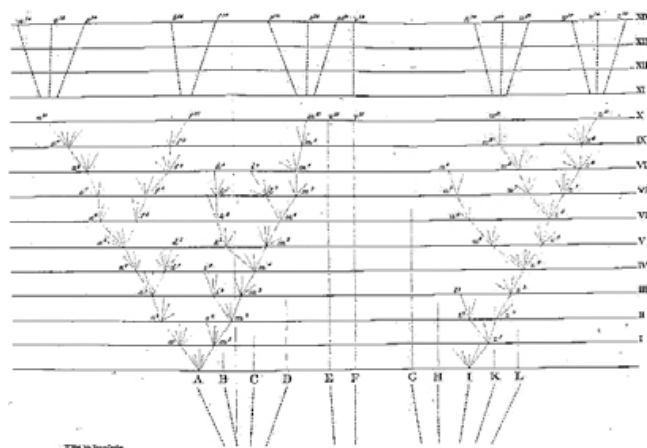


Figure 2: Charles Darwin's family tree (Darwin 1859: illustration in the addendum)

LATERAL TRANSFER IN GENOME EVOLUTION AND LANGUAGE HISTORY

In biology, the family tree model remained the leading evolutionary model for a long time. In linguistics, however, scholars began quite early to question its adequacy to depict the complexity of language history in a realistic manner. As Geisler & List (this volume) emphasize, linguists have long recognized that historical relations between languages are not necessarily *vertical*, i.e. genealogical, but may also be *horizontal*, i.e. non-genealogical, resulting from language contact or lexical borrowing. This finds its reflection in the fact that not long after Schleicher popularized the use of family trees in linguistics, Johannes Schmidt (1843–1901) proposed his *Wave Theory*, as an alternative theory of language evolution. However, lacking the suggestive force of the tree metaphor, Schmidt's *Wave Theory* remained an impalpable concept, as can be seen from the many different attempts in the history of linguistics to visualize it properly (cf. Geisler & List this volume). For many years, linguists would use the family tree while at the same time criticizing its adequacy. Although many scholars followed Schmidt's example and emphasized the inadequacy of the family tree in linguistics, none of the many alternative models that were proposed, be it waves (Schmidt 1872; Hirt 1905), chains, or even animated pictures (Schuchardt 1870 [1900]), gained acceptance among all scholars. For many years, linguists would use the family tree while at the same time criticizing its adequacy.

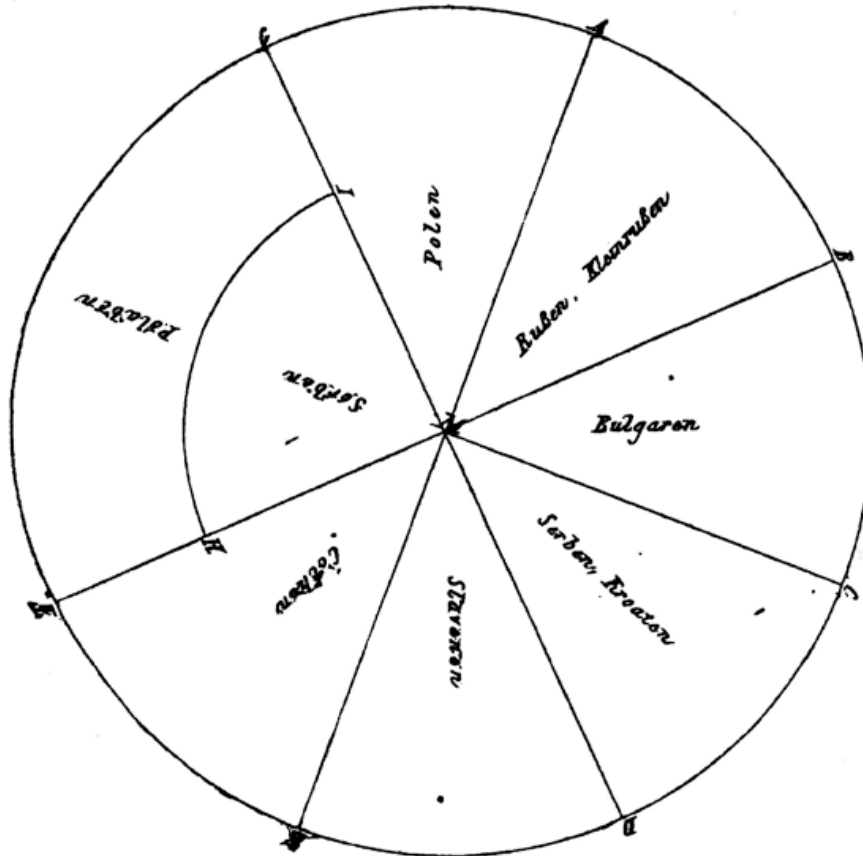


Figure 3: Schmidt's 'Wave Theory' in his own visualization (Schmidt 1875:199)

In contrast to the controversial character of the tree model in linguistics, the bifurcating trees were considered as the only reasonable model to describe species evolution for many decades. While this still holds for macroscopic evolutionary processes, studies in microbiology revealed that the tree model is insufficient for an adequate description of microbial evolution (Dickerson 1980; Doolittle 1999; Ochman 2000; Baptiste et al. 2009). Prokaryotes are capable of acquiring new genetic material from their neighbourhood or directly from the environment and incorporate it into their genomes in a process termed lateral gene transfer (LGT). Gene acquisition by lateral transfer in prokaryotes was first described in the 1950ies (Freeman 1951; Ochiai et al. 1959). The evolutionary implications of LGT bear strong resemblances to the process of borrowing during language evolution (Bryant et al. 2005; Pagel 2009). The development of advanced genome sequencing technologies enabled the investigation of microbial genomes at the DNA level, which led to the realization that LGT plays a major role in shaping microbial genomes (Koonin 2009; Baptiste et al. 2009; Popa and Dagan 2011).

Mechanisms of lateral gene transfer

Known mechanisms of lateral gene transfer include transformation, transduction, conjugation, and gene transfer agents (Thomas and Nielson 2005; Lang and Beatty 2007). Transformation involves the uptake of naked DNA from the environment. DNA uptake is enabled during a competence state that involves 20–50 proteins, including the type IV pillus and type II secretion system proteins (Chen and Dubnau. 2004; Thomas and Nielson 2005). In some species, an effective transformation requires the presence of uptake signal sequences (USSs). These are specific DNA motifs, about 10bp long, that are encoded within the recipient genome in a frequency that is much above that expected by random (Smith et al. 1995). Environmental DNA molecules bearing the USS motif are recognized by specific receptors at the cell surface, imported into the cytoplasm, and can then be readily integrated into the recipient chromosomes, usually via homologous recombination (Chen and Dubnau 2004).

Transduction is DNA acquisition following a phage infection. Phage recognise possible hosts by specific receptors found on the cell surface. Many phages include in their genomes chunks of DNA taken coincidentally from previous hosts. These are transferred to the new host during the integration of the phage genome into the host chromosomes. DNA integration into the host chromosome is generally mediated by the phage-encoded enzymes that specifically integrate the phage into the chromosome of the infected recipient (Thomas and Nielsen 2005; Lindell et al. 2004; Sullivan et al. 2006).

Conjugation is the transfer of DNA via plasmids, a process that is mediated by cell-to-cell junction and a tunnel through which the DNA is transferred. The transferred material is typically a plasmid that can pass through the tunnel during conjugation breaks off. Plasmids can integrate into the recipient chromosomes by homologous recombination that may entail insertion sequences (ISs) or other se-

quences conserved between plasmid and recipient chromosomes that carry the minimal sequence similarity required for homologous recombination (Chen et al. 2005).

Gene transfer agents (GTA) are phage-like DNA-vehicles that are produced by a donor cell and released to the environment. DNA stored in GTAs is imported into the recipient in a process that is similar to transduction. GTAs, unlike phages, are linked to transfer of genomic DNA only and they have no negative effect on the recipient. GTA systems have been documented mainly in oceanic α -proteobacteria, but also in few archaeobacteria and some spirochaetes (Lang et al. 2012; Berglund et al. 2009; Zaho et al. 2010). A recent comparison of GTA-mediated gene transfer rates among various marine habitats revealed particularly high transfer rate in the open ocean, indicating the importance of this transfer mechanism for genome evolution in oceanic α -proteobacteria (Mcdaniel et al. 2010).

An additional transfer mechanism – nanotubes – was discovered recently (Dubey and Ben-Yehuda 2011). These are tubular protrusions composed of membrane components that can bridge between neighboring cells and conduct the transfer of DNA and proteins.

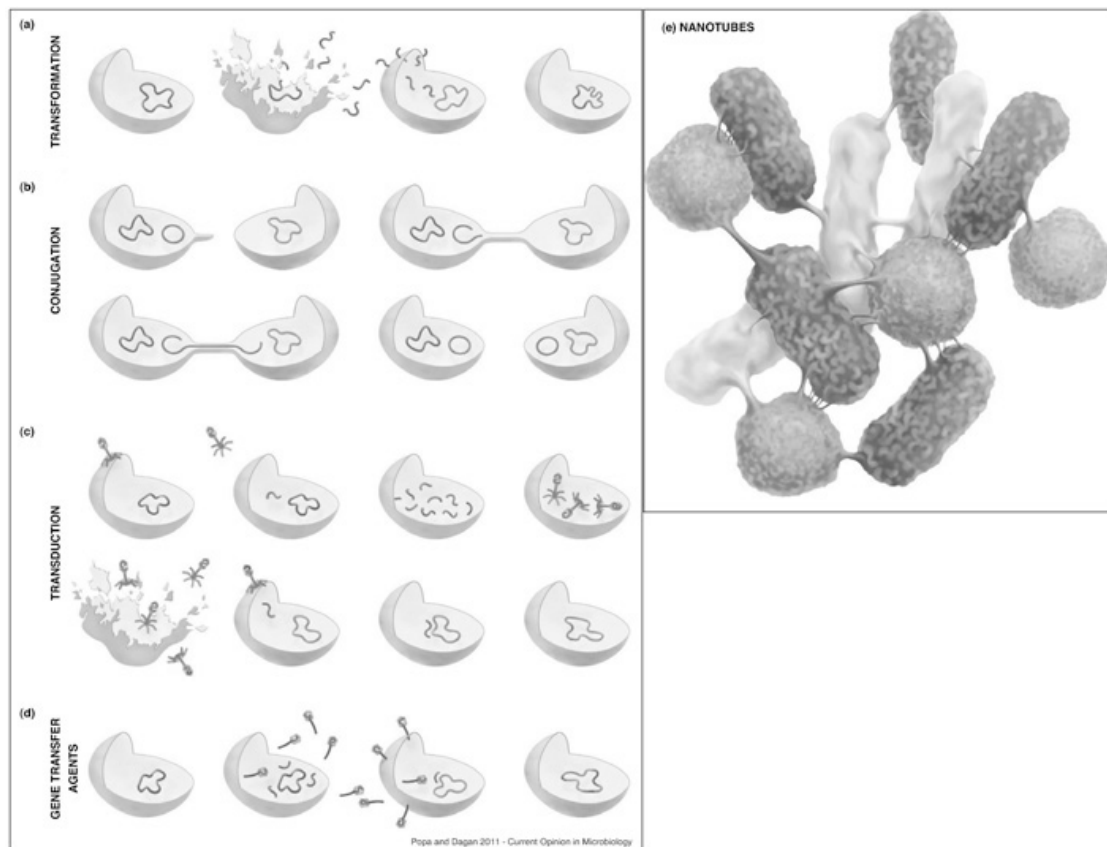


Figure 4: Lateral gene transfer mechanisms

Basic mechanisms of borrowing

Describing the character of languages in all their grammatical, phonetic, and lexical complexity is an extremely difficult task, even when disregarding their history. Consequently, the detailed study of language evolution is often reduced to the study of *lexical change*. The most common unit of the lexicon in a language is the word (or morpheme), a unit which is characterized by its *form*, and its *meaning* (similar to gene sequence and *function* in biology). While the form of a word is directly accessible and can be characterized as a sequence of sound segments, the meaning of normally polysemous words is far more difficult to describe. Like in biology, where protein function cannot be predicted from its sequence alone (in the lack of known homologs), no natural link between form and meaning (function) can be claimed for the linguistic sign. Any connection between form and meaning is only due to convention.

The basic processes of lexical change can be roughly divided into *vertical processes* resulting from semantic change or semantic innovations and *horizontal processes* resulting from borrowing. While vertical processes are *gradual*, horizontal processes are *discrete*, involving a *donor* and a *recipient* language. Evolutionary events affecting a single word can be roughly divided into those that change its form (*sound change*) and those that change its meaning (*semantic change*). Sound change is an overwhelmingly regular process (Hock and Joseph 1995: 241–278). Lexical change is defined as a change in the meaning of a sign compared to its ancestor while a change in the form resulting from regular sound change processes is disregarded (Gevaudan 2007: 14f). In a *direct transfer* both the form and meaning of a word are transferred as a whole from the donor to the recipient language. During *semantic transfer* (or *semantic borrowing*) a word in the donor language is reproduced in the recipient language by expanding the meaning of a given word in the recipient language to match the form-meaning unity in the donor language (*semantic transfer*, *semantic borrowing*, cf. Weinreich 1953: 48). For example, the standard Chinese *kāfēi* “coffee” was directly transferred from English, as is also evident from the similar pronunciation of the words. The standard Chinese *diàn* “electricity”, on the other hand, has been indirectly transferred by extending the word’s original meaning “lightning” (see Figure 5).

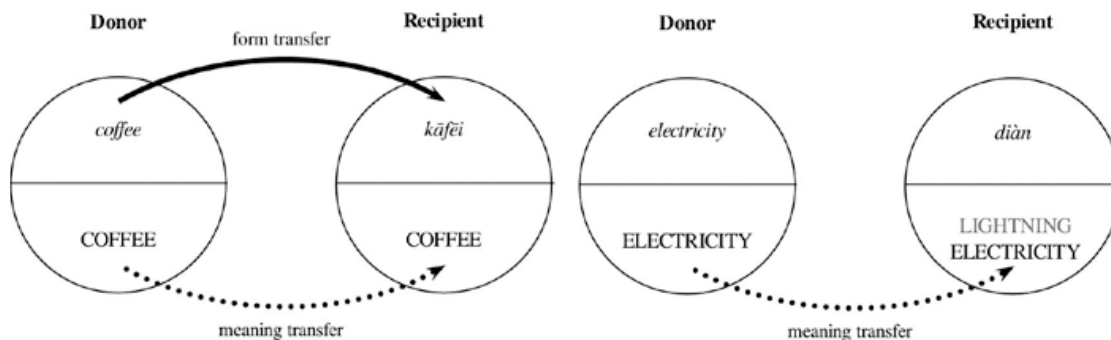


Figure 5: Direct transfer and reproduction

SIMILARITIES AND DIFFERENCES IN LANGUAGE AND GENOME EVOLUTION

The most striking difference between languages and genomes, as Geisler and List (this volume) point out, lies in the way they manifest themselves: genome evolution takes place in a materialistic world, while language history does not. Genes have a substance, words do not, and while the function of a gene is a product of its sequence, any connection between form and function of the linguistic sign is strictly arbitrary. These differences are reflected in the mechanisms that drive lateral gene transfer and lexical borrowing. The most striking of these differences is that the transfer of linguistic material is not restricted to the direct transfer of form. Lateral gene transfer only applies to the exchange of genetic material, i.e. the exchange of form as a vehicle of function. In contrary, lexical borrowing between languages can involve the transfer of both form and function, or the transfer of functions alone.

Lateral transfer frequency during genome evolution

Several experiments have been conducted in order to quantify the frequency of LGT in nature. For example, Babic et al. (2008) tested the success rate of gene acquisition by conjugation in *Escherichia coli*. Using a plasmid encoding a gene for fluorescence protein (YFP) they quantified the odds for a successful integration of plasmid genes into the recipient genome. They found that in 96% of the population the YFP gene was integrated into the chromosome and inherited to the next generation. The percolation of an acquired DNA within the population can be extremely fast in *Bacillus subtilis* where the cells are arranged in chains. Tracking the spread of an integrative and conjugative element (ICE) encoding a gene for green fluorescence protein (GFP) under the microscope showed that in 43 (81%) out of 53 cases a recipient cell turned into a donor and transconjugated the ICE to the next cell in line, often within 30 minutes (Babic et al. 2011).

Lateral gene transfer via transduction takes place during a phage infection. Hence gene acquisition by this transfer mechanism depends on the survival of the recipient. In a recent study Kenzaka et al. (2010) quantified the survival rate of phage infected enteric bacteria as 20% of the population. These surviving bacteria may acquire DNA from previous hosts of the attacking phage.

We know that LGT occurs in the laboratory, the issue is how often it occurs in the wild and how important it is during evolution. Phylogenetic reconstruction of microbial genes reveals that LGT plays a major role in shaping microbial genomes (Mirkin et al. 2003; Kunin et al. 2005; Dagan and Martin 2007; Halary et al. 2010; Kloesges et al. 2011). In a pioneering study, Lawrence and Ochman (1998) identified all *E. coli* genes that were acquired since its divergence from the *Salmonella* lineage by their aberrant codon usage. They estimated that 755 (18%) of the 4,288 genes in *E. coli* strain MG1655 were laterally acquired over a time period of about 14 million years (Myr) and estimated the LGT rate as 16Kb/1Myr per lineage (Lawrence and Ochman 1998). Using gene distribution patterns across 329 proteobacte-

rial genomes, Kloesges et al. (2011) recently estimated that at least 75% of the protein families have been affected by LGT during evolution. Gene transfer rate in those families is on average 1.9 events per protein family per lifespan (Kloesges et al. 2011). Similar estimates were found in phylogenetic analyses of broader taxonomic samples (Mirkin et al. 2003; Kunin et al. 2005; Dagan and Martin 2007).

The impact of LGT during genome evolution can be estimated either by the proportion of recently transferred genes whose unusual base composition and codon usage still bears the marks of acquired DNA (Lawrence and Ochman 1998; Garcia-Vallve et al. 2000; Nakamura et al. 2004) or by phylogenetic analysis of individual genes including recent and ancient LGTs alike (e.g. Zhaxybayeva et al. 2006; Beiko et al. 2005; Puigbò et al. 2010; Chan et al. 2011). A survey of genes having aberrant nucleotide composition within proteobacterial genomes revealed that $21\pm 9\%$ of the genes in those genomes comprises recent acquisitions (Kloesges et al. 2011). Gene distribution patterns across the same species sample suggest that, on average, $74\pm 11\%$ of the genes in each genome have been laterally transferred at least once during evolution (Kloesges et al. 2011).

Lateral transfer frequency during language history

Borrowing frequency may vary dramatically during language evolution, depending on many different factors such as the sociocultural situation in which the respective language is used, the geographical distance of the language to other languages, or the prestige of specific language varieties within a given speech community. Borrowed vocabulary can affect only small parts of the lexicon of a given language (such as specific terms for cultural items), or alternatively result in a situation where large parts of the language lexicon are acquired or replaced and can be traced back to a donor language.

In the recently published World Loanword Database (WOLD, Haspelmath and Tadmor 2009) the frequency of direct borrowing events in a sample of 1460 glosses that were translated into 41 different languages was investigated. Borrowing rates in the database vary greatly, ranging from 1 % (Mandarin Chinese) to 62 % (Selice Romani) with an average of 25 % and a standard deviation of 13 % (Tadmor 2009; Figure 6)

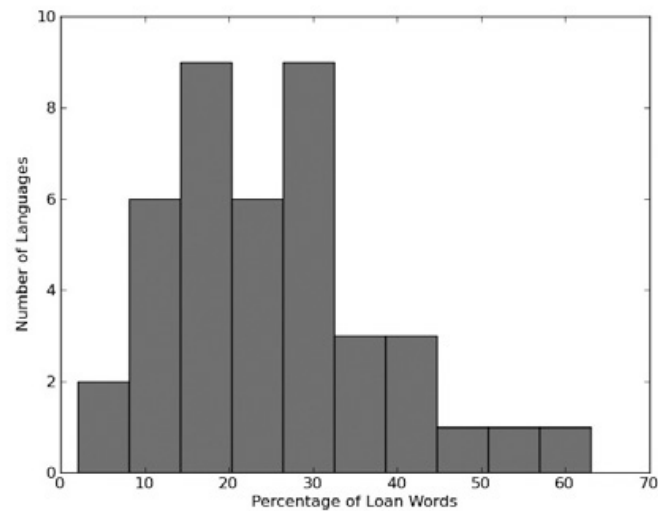


Figure 6: Histogram of borrowing frequencies in the WOLD database

PHYLOGENOMIC NETWORKS

Considering the frequency of lateral transfer events during the evolution of genomes and languages leads to the realisation that the tree model is missing a substantial portion of the evolutionary history in these domains. Studying the evolutionary dynamics of genomes and languages in detail requires alternative, more complex, models that allow to incorporate both vertical and horizontal transfers. Networks approach provides a more realistic model of microbial and language evolution than trees because they allow the reconstruction of non tree-like events such as recombination, gene fusion, and lateral gene transfer (Huson and Scornavacca 2011; Dagan 2011). Representing genealogical relations using a network is not new and has been documented even before Darwin's species tree was popularized (Ragan 2009). These earlier examples however focus on illustrating complex evolutionary relationship. The advance in statistical methods to analyse network properties enables the application of networks to evolutionary studies in a much more quantitative way (Dagan 2011).

A network constitutes a set of entities and the pairwise relations among them. The entities are termed *nodes* (or *vertices*) are connected by *edges* representation relationship (Newman 2010). Phylogenomic networks comprise completely sequenced genomes as nodes that are connected by edges of phylogenetic relations (Dagan 2011). Phylogenomic networks can be reconstructed from shared gene content (Dagan et al. 2008; Halary et al. 2010), shared sequence similarity (Lima-Mendez et al. 2008), or phylogenetic trees (Beiko et al. 2005; Popa et al. 2011). A more detailed phylogenomic network is the *directed network of lateral gene transfer* (dLGT) in which the nodes correspond to species or their ancestors and the directed edges represent recent lateral transfer events containing direction from donor to recipient as additional information (Popa et al. 2011).

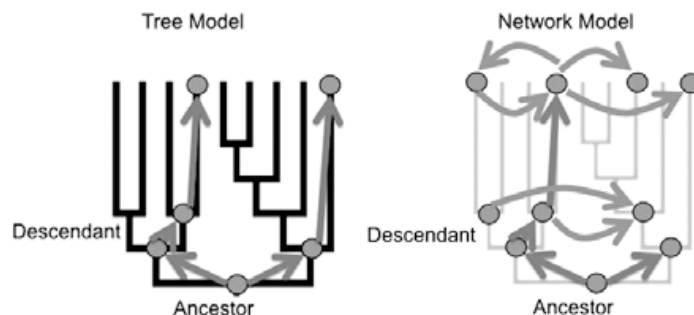


Figure 7: 'Tree Model' and 'Network Model'

Phylogenetic networks of languages

Linguists have long been aware of the problems that lexical borrowing poses to the tree model and tried to find more sophisticated ways to include the lateral component of language evolution in the evolutionary process representation. Given the need to model both vertical and horizontal processes, linguists would naturally turn to networks as a format to represent language evolution. Ignoring spurious hints to horizontal branches introduced into language trees that can be found in the literature rather early (Schuchard 1870 [1900]), the first explicit network approach can be found in a study by Bonfante (1931) where the complex relations between the Indo-European languages led the author to deny the possibility of true genealogical tree in this language family and propose a network model instead (see Figure 5 in Geisler & List this volume). Unfortunately, the use of networks by Bonfante and later similar studies (Southworth 1964, Anttila 1972) remained a mere visualization of the scholars' intuitions regarding patterns in the data and did not enable further insights regarding language evolution.

More recent attempts to reconstruct networks of language phylogeny include the applications of reticulated trees and split networks, which were initially developed in order to study genome evolution (Bryant 2005, McMahon 2005, Hamed and Wang 2006). While these methods can reveal the extent of non tree-like evolutionary dynamics, none of them can be used to estimate borrowing frequency. Thus, although based on quantitative data, split networks still remain a visualization tool.

Minimal lateral networks (MLN; Dagan et al. 2008), developed originally to study microbial genome evolution, enable an automatic inference of borrowing events in linguistic datasets and their visualization in a network form (Nelson-Sathi et al. 2011). The method is based on the construction of a phylogenetic network model in which presumed cognate words, i.e. words that go back to a common ancestor, are mapped on to a reference tree for the languages being analysed. Branching patterns in the data that are incompatible with the reference tree are explained by means of different borrowing models. These models allow for an increasing amount of borrowing events by which the patchiness of the data in comparison with the reference tree can be explained. Based on the assumption that the number of words that are used to express a certain set of concepts is approximately the same in all languages and throughout all times, a model is chosen which mini-

mizes the difference in the average word inventories between the ancestor language and its descendants.

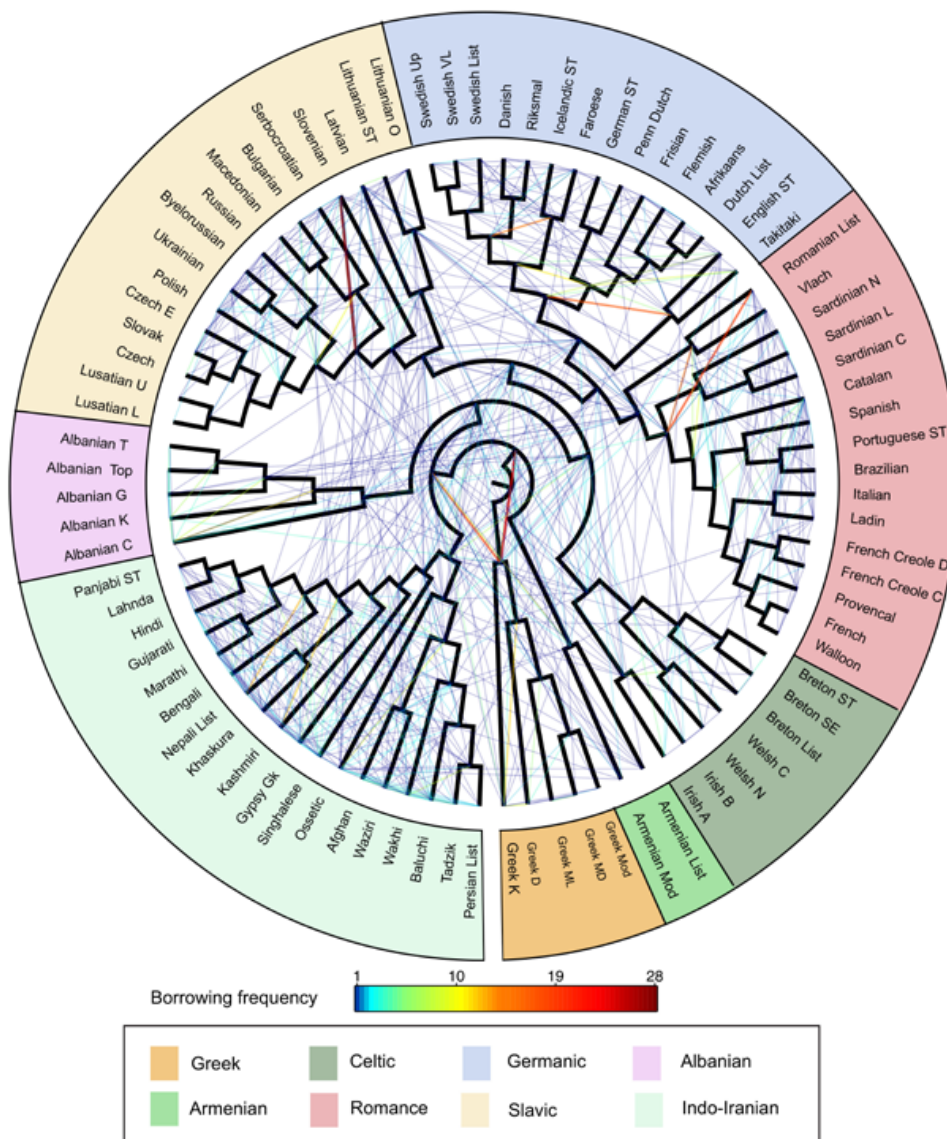


Figure 8: Minimal lateral network of Indo-European languages

Figure 8 shows a Minimal Lateral Network (MLN) of a dataset of 84 Indo-European languages (Dyen et al. 1992) representing both the vertical as well as the lateral components of language evolution reconstructed by the method. Nodes in this network represent contemporary (external) and ancestral (internal) languages. The ancestral languages correspond to the nodes in the reference tree. The vertices are connected either by branches of the reference tree, representing vertical inheritance, or by lateral edges, representing inferred borrowing events whose frequency corresponds to the edge width.

The method has several advantages over previously proposed ones: Since it infers concrete borrowing events, the results are transparent and can be directly tested against the data. Furthermore, additional analyses can be applied to the MLN

addressing various questions such as what is the borrowing frequency observed in the data or which trends and barriers exist for borrowing dynamics during language evolution.

The benefits of network approaches in biology and linguistics

Phylogenetic networks – in contrast to phylogenetic trees – have many advantages when studying genome evolution or language history. They are more informative than simple family trees, since they do not ignore horizontal relations between genomes and languages. In contrast to the different Wave models, which were presented as alternatives to the tree model in historical linguistics (see Geisler and List this volume), phylogenetic networks do not result in a static visualization of relations between taxa, but provide a dynamic model of evolution in both linguistics and biology.

TRENDS AND BARRIERS

For a long time, the lateral components of genome evolution and language history have been ignored by biologists and linguists. Reticulated evolutionary events such as lateral transfer are often seen as irregular, chaotic events that blur the “real” phylogeny in both fields. The ability to include the lateral component of genome evolution and language history as an integral part of phylogenetic studies is, however, expected to promote our understanding of the trends and the barriers that underlay LGT and lexical borrowing dynamics. Experimental work shows that gene acquisition by LGT among prokaryotes is frequent and that the percolation of acquired DNA among populations and across generations is rapid. Phylogenomic analyses reveal that LGT has a substantial impact on long-term genome evolution, supplying a mechanism for natural variation that is specific for the prokaryotic domains and allows their adaptation in dynamic environments. Prokaryote genome evolution comprises thus vertical (tree-like) and lateral (network-like) components. At the same time, different types of barriers to LGT on the genomic, species, and habitat levels are becoming increasingly apparent (Popa and Dagan 2011).

Trends and barriers to lateral gene transfer

The dLGT network reveals clusters of densely connected donors and recipients that are very similar in their genomic nucleotide content (GC content) (Popa and Dagan 2011). The difference in genomic GC content between donors and recipients is <5% for most (86%) of connected pairs (Popa et al 2011). Furthermore, donor-recipient genome sequence similarity and LGT frequency are positively correlated ($r_s = 0.55$, $P \ll 0.01$) (Popa et al. 2011). This suggests that LGT is more frequent among closely related species, having similar genomes, while LGT between distantly re-

lated species is more rare bring out the existence of a donor-recipient *similarity barrier*.

Microbes tend to delete non-functional or otherwise unneeded DNA from their genomes (Moran et al 2009, Burke and Moran 2011). Therefore, the fixation of the acquired DNA within the genome is highly dependent on its functionality or utility to the recipient under selectable environmental conditions (Hao and Golding 2006; Pal et al. 2005; Zhaxybayeva and Doolittle 2011). Most laterally transferred genes perform *metabolic functions*, e.g. they are responsible for harvesting energy or for the construction of cell components like amino- or nucleic acids, while the transfer of genes performing information processing (including replication, transcription, and translation) is rare (Popa et al 2011, Jain et al. 1999, Coscolla et al 2011). According to the complexity hypothesis (Jain et al. 1999), the scarcity of lateral transfer of information processing genes is attributed to their role in complex structures. Proteins that function in a complex structure, for example ribosomal proteins, are adapted to their common function. An LGT event that leads to replacement of such a gene with a less adapted homolog will result in a ‘squeaking wheel’ within the complex and reduced fitness of the recipient (Jain et al. 1999). This suggests that there is a *functional barrier* for LGT. Genomic fragments whose cloning into *E. coli* is lethal are suspects for encoding proteins whose acquisition in *E. coli* is extremely disadvantageous (Sorek et al. 2007). An extensive dataset of lethal fragments collected during genome sequencing projects of 79 diverse species showed that these fragments typically encode for single copy genes. The integration of an additional gene copy into the *E. coli* genome resulted in an elevated protein production that was lethal to the cell (Sorek et al. 2007) suggesting protein dosage as another functional barrier to LGT.

The physical distance between the donor and recipient in the LGT event depends upon the LGT mechanism (Popa and Dagan 2011). In transformation the distance between the donor and recipient depends upon the raw DNA stability within the environment (Majewski J, 2001). Conjugation requires that the donor and recipient will be close enough for the formation of the conjugation tunnel. Transduction is considered as the longest range LGT mechanism because it entails phage mobility (Majewski J, 2001). This suggests that most transfers should occur within habitats. The dLGT network reveals that indeed most (74%) of the detected LGT in the network occur between donors and recipients residing in the same habitat (Popa and Dagan 2011), indicating the presence of an *ecological barrier* to LGT. A network of shared transposases among 774 microbial genomes supplies further support for the rarity of interhabitat gene transfers (Hooper et al. 2009). Halary et al. (2010) reconstructed a network of shared protein families among various genetic entities including microbial chromosomes, plasmids, and phage genomes. A comparison of network properties between plasmids and phage genomes revealed that plasmids are more frequently connected within the network in comparison to phages. From this they concluded that conjugation is more frequent than transduction in nature (Halary et al. 2010).

Trends and barriers to lexical borrowing

As in biology, there are certain barriers for horizontal transfer during language evolution. Since the sound systems of languages can be very different, the pronunciation difficulty of borrowed word within the recipient population may vary. In cases where the difference in the sound systems of donor and recipient languages is sufficiently large, direct borrowing will occur less frequently, resulting in a *similarity barrier*. For example, although the daily life in China is heavily influenced by Western culture, English words for Western concepts are rarely directly transferred into the Chinese language, but rather reproduced due to the large differences in the sound systems of English and Chinese. This is probably the reason why the borrowing capacity of Mandarin Chinese is the lowest in the sample of WOLD with only 1.2% direct borrowings out of a total of 2042 words (Wiebusch 2009): Although in Standard Chinese there are many terms which have been coined under the influence of foreign examples, these words are expressed in a seemingly genuine manner. For example, the Standard Chinese word for “boomerang” is *fēiqùláiqì*, which literally can be translated as “there-and-back-flying device”.

Borrowing events will be less frequent between geographically distant speech communities, resulting in a *spatial barrier*. The spatial barrier is closely connected with what one might call a *socio-cultural* or *socio-political barrier* for lexical borrowing: Due to social, cultural, or political reasons a given language variety may either be promoted or marginalized by the ones who speak it, resulting in a high or low borrowing rates (Tadmor 2009).

Furthermore, given that most borrowed words are due to the lack of certain words for certain concepts in the recipient language, that are present in the donor language, borrowing heavily depends on the meaning of the items being borrowed. While the exchange of innovations between different communities is often also accompanied by the exchange of lexical items, words denoting basic concepts that are essential for human life are less likely to be exchanged, resulting in a *functional barrier* (Hock and Joseph 2009).

OUTLOOK

Given the fact that language history and genome evolution take place in very different domains, it is not surprising to find many differences between both processes, especially when dealing with the details of the mechanisms and their explanations. From a more abstract perspective, there are, however, many interesting similarities between language history and genome evolution that are revealed when comparing the trends and the barriers to lateral transfer in the linguistic and the biological domains.

Understanding how languages and genomes change, how they eventually split, separate, and diverge, is a challenging problem in evolutionary biology and historical linguistics. Many different methods have been proposed so far to study these processes in detail. During the last two decades, linguists have especially focused

on quantifying the traditional qualitative methods. As a result, many new approaches to language phylogenetic reconstruction have been proposed, leading to a better understanding of the genealogical processes that led to the diversification of different language families. However, because language change is not only based on the modification of inherited items but is also driven by the direct or indirect transfer of linguistic units, phylogenetic trees do not tell us the true story about the history of languages, but provide only a reduced version which may be often misleading. The same holds for microbial evolution where the high frequency of LGT renders the tree model insufficient. In both disciplines network approaches can assist to uncover on reticulated evolutionary events that were previously ignored.

REFERENCES

- Babic, A.; Berkmen, M.B.; Lee, C.A. & A.D. Grossman (2011) 'Efficient Gene Transfer in Bacterial Cell Chains', *mBio* 2, e00027–11–e00027–11.
- Babic, A.; Lindner, A.B.; Vulic, M.; Stewart, E.J. & M. Radman (2008) 'Direct Visualization of Horizontal Gene Transfer', *Science* 319: 1533–6.
- Baptiste, E.; O'Malley, M. & R. Beiko (2009) 'Prokaryotic evolution and the tree of life are two different things', *Biol Direct* PMID: 19788731.
- Beiko, R.G.; Harlow, T.J. & M.A. Ragan (2005) 'Highways of gene sharing in prokaryotes', *Proc Natl Acad Sci USA* 102: 14332–7.
- Berglund, E.C.; Frank, A.C.; Calteau, A.; Vinnere Pettersson, O.; Granberg, F.; Eriksson, A.-S.; Näslund, K.; Holmberg, M.; Lindroos, H. & S.G.E. Andersson (2009) 'Run-Off Replication of Host-Adaptability Genes Is Associated with Gene Transfer Agents in the Genome of Mouse-Infecting *Bartonella grahamii*', *PLoS Genet* 5, e1000546.
- Bonfante, G. (1931) 'I dialetti indoeuropei', in *Annali del R. Istituto Orientale di Napoli* 4: 69–185.
- Burke GR, Moran NA (2011) 'Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids', *Genome Biol Evol* 3: 195–208.
- Bryant, D.; Filimon, F. & R.D. Gray (2005) 'Untangling our past: Languages, Trees, Splits and Networks', in R. Mace, C.J. Holden & S. Shennan (eds), *The evolution of cultural diversity: A phylogenetic approach* (London: UCL Press): 69–85.
- Cannon, W.F. (1960) 'The uniformitarian-catastrophist debate', *Isis* 51(1): 38–55.
- Chan, C.X.; Beiko, R.G. & M.A. Ragan (2011) 'Lateral Transfer of Genes and Gene Fragments in *Staphylococcus* Extends beyond Mobile Elements', *Journal of Bacteriology* 193: 3964–77.
- Chen I, & D. Dubnau (2004) 'DNA uptake during bacterial transformation', *Nat Rev Microbiol* 2: 241–9.
- Chen, I.; Christie, P. & D. Dubnau (2005) The ins and outs of DNA transfer in bacteria. *Science* 2(310/5753): 1456–60.
- Coscolla M.; Comas I. & F. Gonza lez-Candelas (2011) 'Quantifying nonvertical inheritance in the evolution of *Legionella pneumophila*', *Mol Biol Evol* 28: 985–1001.
- Dagan, T. (2011) 'Phylogenomic networks', *Trends in Microbiology* 19: 483–91.
- Dagan, T. & W. Martin (2007) 'Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution', *Proc Natl Acad Sci USA* 104: 870–5.
- Dagan, T.; Artzy-Randrup, Y. & W. Martin (2008) 'Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution', *Proc Natl Acad Sci USA* 10039–44.
- Darwin, C. (1859) *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life* (London: John Murray).
- Dickerson, R. (1980) 'Evolution and gene transfer in purple photosynthetic bacteria', *Nature* 283(5743): 210–2.

- Doolittle, W. Ford (1999) 'Phylogenetic classification and the universal tree', *Science* 284(5423): 2124–9.
- Dubey, G.P. & S. Ben-Yehuda (2011) 'Intercellular Nanotubes Mediate Bacterial Communication', *Cell* 144: 590–600.
- Dyen, I.; Kruskal, J.B. & P. Black (1992) 'An Indoeuropean classification: A lexicostatistical experiment', *Trans. Am. Philos. Soc.* 82: 1–132.
- Freeman, V.J. (1951) 'Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*', *Journal of Bacteriology* 61: 675.
- Garcia-Vallvé, S.; Romeu, A. & J. Palau (2000) 'Horizontal gene transfer in bacterial and archaeal complete genomes', *Genome Research* 10: 1719–25.
- Gèvaudan, P. (2007) *Typologie des lexikalischen Wandels. Bedeutungswandel, Wortbildung und Entlehnung am Beispiel der romanischen Sprachen* (Tübingen: Stauffenburg).
- Gogarten J.P.; Doolittle, W.F. & J.G. Lawrence (2002) 'Prokaryotic evolution in light of gene transfer', *Molecular Biology and Evolution* 19(12): 2226–38.
- Halary, S.; Leigh, J.W.; Cheaib, B.; Lopez, P. & E. Bapteste (2010) 'Network analyses structure genetic diversity in independent genetic worlds', *Proc Natl Acad Sci USA* 107: 127–32.
- Hao, W. & G.B. Golding (2006) 'The fate of laterally transferred genes: life in the fast lane to adaptation or death', *Genome Res*, 16: 636–643.
- Hamed, M.B. & F. Wang (2006) 'Stuck in the forest: Trees, networks and Chinese dialects', *Diachronica* 23: 29–60(32).
- Haspelmath, M. & U. Tadmor (eds) (2009) *Loanwords in the World's Languages. A Comparative Handbook* (Berlin/New York: de Gruyter).
- Hirt, H. (1905) *Die Indogermanen. Ihre Verbreitung, ihre Urheimat und ihre Kultur*, Vol. 1 (Strassburg: Trübner).
- Hock, H.H. & B.D. Joseph (1995 [2009]) *Language history, language change and language relationship. An introduction to historical and comparative linguistics*, 2nd ed. (Berlin/New York: de Gruyter).
- Holden, C.J. & S. Shennan (2005) *The evolution of cultural diversity* (London: UCL Press): 67–84.
- Huson, D.H. & C. Scornavacca (2011) 'A Survey of Combinatorial Methods for Phylogenetic Networks', *Genome Biology and Evolution* 3: 23–35.
- Jain, R.; Rivera, M.C. & J.A. Lake (1999) 'Horizontal gene transfer among genomes: the complexity hypothesis', *Proc Natl Acad Sci USA* 96: 3801–6.
- Kenzaka, T., Tani, K. & M. Nasu (2010) 'High-frequency phage-mediated gene transfer in freshwater environments determined at single-cell level', *The ISME Journal* 4: 648–59.
- Kloesges, T., O. Popa, W. Martin & T. Dagan (2011) 'Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths', *Molecular Biology and Evolution* 28: 1057–74.
- Koonin, E.V. (2009) 'Darwinian evolution in the light of genomics', *Nucleic Acids Research* 37: 1011–34.
- Kunin, V. (2005) 'The net of life: Reconstructing the microbial phylogenetic network', *Genome Research* 15: 954–9.
- Lang, A.S. & J.T. Beatty (2007) 'Importance of widespread gene transfer agent genes in alpha-proteobacteria', *Trends in Microbiology* 15: 54–62.
- Lang, A.S.; Zhaxybayeva, O. & J.T. Beatty (2012) 'Gene transfer agents: phage-like elements of genetic exchange', *Nat Rev Micro* 10: 472–82.
- Lawrence, J.G. & H. Ochman (1998) 'Molecular archaeology of the *Escherichia coli* genome', *Proc Natl Acad Sci USA* 95: 9413–7.
- Lima-Mendez, G., J. Van Helden, A. Toussaint & R. Leplae (2008) 'Reticulate Representation of Evolutionary and Functional Relationships between Phage Genomes', *Molecular Biology and Evolution* 25: 762–77.
- Majewski, J. & F.M. Cohan (1999) 'DNA sequence similarity requirements for interspecific recombination in *Bacillus*', *Genetics* 153: 1525–33.

- Martin, W. (1999) 'Mosaic bacterial chromosomes: a challenge en route to a tree of genomes', *Bioessays* 21: 99–104.
- McDaniel, L.D.; Young, E.; Delaney, J.; Ruhnu, F.; Ritchie, K.B. & J.H. Paul (2010) 'High Frequency of Horizontal Gene Transfer in the Oceans', *Science* 330: 50.
- McMahon, A. & R. McMahon (2005) *Language classification by numbers* (Oxford: Oxford University Press).
- Mirkin, B.G.; Fenner, T.I.; Galperin, M.Y. & E.V. Koonin (2003) 'Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes', *BMC Evol Biol* 3: 2.
- Moran, N.A.; McLaughlin, H.J. & R. Sorek (2009) 'The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria', *Science* 323: 379–82.
- Nakamura, Y.; Itoh, T.; Matsuda, H. & T. Gojobori (2004) 'Biased biological functions of horizontally transferred genes in prokaryotic genomes', *Nat Genet* 36: 760–6.
- Nakhleh, L.; Ringe, D. & T. Warnow (2005) 'Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages', *Language* 81.2: 382–420.
- Navarre, W.W.; Porwollik, S.; Wang, Y.; McClelland, M.; Rosen, H.; Libby, S.J. & F.C. Fang (2006) 'Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*', *Science* 313: 236–8.
- Nelson-Sathi, S.; List, J.-M.; Geisler, H.; Fangerau, H.; Gray, R.D.; Martin, W. & T. Dagan (2011) 'Networks uncover hidden lexical borrowing in Indo-European language evolution', *Proceedings of the Royal Society B. Biological Sciences* 278.1713: 1794–803.
- Norman, A., Hansen, L.H. & S.J. Sørensen (2009) 'Conjugative plasmids: vessels of the communal gene pool', *Philosophical Transactions of the Royal Society B, Biological Sciences* 364: 2275–89.
- Ochman, H.; Lawrence, J.G. & E.A. Groisman (2000) 'Lateral gene transfer and the nature of bacterial innovation', *Nature* 405: 299–304.
- Pagel, M. (2009) 'Human language as a culturally transmitted replicator', *Nature Reviews Genetics* 10: 405–15.
- Pal, C.; Papp, B. & M.J. Lercher (2005) 'Adaptive evolution of bacterial metabolic networks by horizontal gene transfer', *Nat Genet* 37: 1372–5.
- Popa, O. & T. Dagan (2011) 'Trends and barriers to lateral gene transfer in prokaryotes', *Current Opinion in Microbiology* 14(5): 615–23.
- Popa, O.; Hazkani-Covo, E.; Landan, G.; Martin, W. & T. Dagan (2011) 'Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes', *Genome Research* 21: 599–609.
- Puigbo, P.; Wolf, Y.I. & E.V. Koonin (2010) 'The Tree and Net Components of Prokaryote Evolution', *Genome Biology and Evolution* 2: 745–56.
- Ragan, M.A. (2009) 'Trees and networks before and after Darwin', *Biol Direct* 4: 43.
- Schleicher, A. (1853a) 'Die ersten Spaltungen des indogermanischen Urvolkes', *Allgemeine Monatsschrift für Wissenschaft und Literatur* Sept 1853: 786–87.
- Schleicher, A. (1853b) 'O jazyku litevském, zvláště na slovanský. Čteno v posezení sekci filologické král', *České Společnosti Nauk dne 6. června 1853*, *Časopis Českého Museum* 27: 320–34.
- Schleicher, A. (1863) *Die Darwinsche Theorie und die Sprachwissenschaft. Offenes Sendschreiben an Herrn Dr. Ernst Haeckel* (Weimar: Böhlau).
- Schmidt, J. (1872) *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen* (Weimar: Böhlau).
- Schuchardt, H. (1900[1870]) *Über die Klassifikation der romanischen Mundarten, Probe-Vorlesung gehalten zu Leipzig am 30. April 1870* (Graz: Styria).
- Smith, H.O.; Tomb, J.F.; Dougherty, B.A.; Fleischmann R.D. & J.C. Venter (1995) 'Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome', *Science* 269: 538–540.
- Sorek, R.; Zhu, Y.; Creevey, C.; Francino, M. & P. Bork (2007) 'Genome-wide experimental determination of barriers to horizontal gene transfer', *Science* 318: 1449–52.

- Southworth, F.C. (1964) 'Family-tree diagrams', *Language* 40.4: 557–65.
- Sullivan, M.B.; Lindell, D.; Lee, J.A.; Thompson, L.R.; Bielawski, J.P. & S.W. Chisholm (2006) 'Prevalence and Evolution of Core Photosystem II Genes in Marine Cyanobacterial Viruses and Their Hosts', *PLoS Biol* 4(8): e234.
- Sutrop, U. (1999) 'Diskussionsbeiträge zur Stammbaumtheorie', *Fenno-Ugristica* 22: 223–51.
- Tadmor, U. (2009) 'Loanwords in the world's languages. Findings and results', in M. Haspelmath & U. Tadmor (eds), *Loanwords in the world's languages. A comparative handbook* (Berlin/New York: de Gruyter): 55–75.
- Thomas, C.M. & K.M. Nielsen (2005) 'Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria', *Nat Rev Micro* 3: 711–21.
- Weinreich, U. (1953 [1974]) *Languages in contact*, With a preface by André Martinet, 8th ed (The Hague / Paris: Mouton).
- Wells, R.S. (1973) 'Lexicostatistics in the regency period', in *Lexicostatistics in Genetic Linguistics*, Proceedings of the Yale Conference, 3./4. April 1971 (Yale University): 118–21.
- Wiebusch, Thekla (2009) 'Mandarin Chinese vocabulary', M. Haspelmath & U. Tadmor (eds), *World Loanword Database* (Munich: Max Planck Digital Library). URL: <http://wold.living-sources.org>.
- Zhao, Y., Wang, K., Ackermann, H.-W., Halden, R.U., Jiao, N., and Chen, F. (2010). Searching for a "Hidden" Prophage in a Marine Bacterium. *Applied and Environmental Microbiology* 76, 589–595.
- Zhaxybayeva, O.; Gogarten, J.P.; Charlebois, R.L.; Doolittle, W.F. & R.T. Papke (2006) 'Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events', *Genome Res* 16: 1099–1108.
- Zhaxybayeva, O. & W.F. Doolittle (2011) 'Lateral gene transfer', *Curr Biol* 21: R242-R246.