

# A Comprehensive Functional Analysis of Ancestral Human Signal Peptides

Matthias Gralle\*<sup>†1</sup> and Svante Pääbo<sup>1</sup>

<sup>1</sup>Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

<sup>†</sup>Present address: Instituto de Bioquímica Médica, Universidade Federal do Rio de Janeiro, Brasil.

\*Corresponding author: E-mail: gralle@bioqmed.ufrj.br.

Associate editor: Carlos Bustamante

## Abstract

With the sequencing of the Neandertal genome, it has become possible to identify amino acid substitutions that occurred on the human lineage since its separation from the Neandertal lineage. Conceptually, it will therefore be possible to functionally analyze all such amino acid substitutions in the future. Here, we analyze the function of substitutions that occurred during recent human evolution in N-terminal signal peptides. We develop a high-throughput flow cytometry-based assay to analyze signal peptide efficiency as the ratio of surface to total reporter protein per live cell. Such ratios differed significantly among signal peptides derived from different human genes. However, no modern human signal peptide differed significantly from its ancestral counterpart, an observation compatible with the predictions of the neutral theory of molecular evolution.

**Key words:** neandertal, human evolution, membrane protein, neutral theory, positive selection.

The identification of potentially important evolutionary changes in the genome is generally done by analyzing patterns of variation within and between species. Such analyses can identify genes or other genomic regions that under evolutionary models have a certain statistical likelihood of having been affected by positive selection. However, the ultimate tests for such genes are functional analyses that demonstrate that the ancestral and derived alleles differ in function (Sabeti et al. 2006). In humans, several screens for positive selection based on genome-wide variation within the species as well as differences to the chimpanzee and other primates have been performed (Sabeti et al. 2006; Akey 2009). Several cases of intraspecific recent or ongoing positive selection of alleles have been described, and these alleles have in some cases been functionally characterized. However, evolutionary changes that are fixed in current humans, but occurred after the human evolutionary lineage diverged from the chimpanzee lineage, are more difficult to identify. This is due to the fact that tests based on intraspecific variation only detect positive selection that occurred recently (Sabeti et al. 2006), whereas tests based on between-species comparisons have little power because of the relatively low number of nucleotide differences between ape and human genomes.

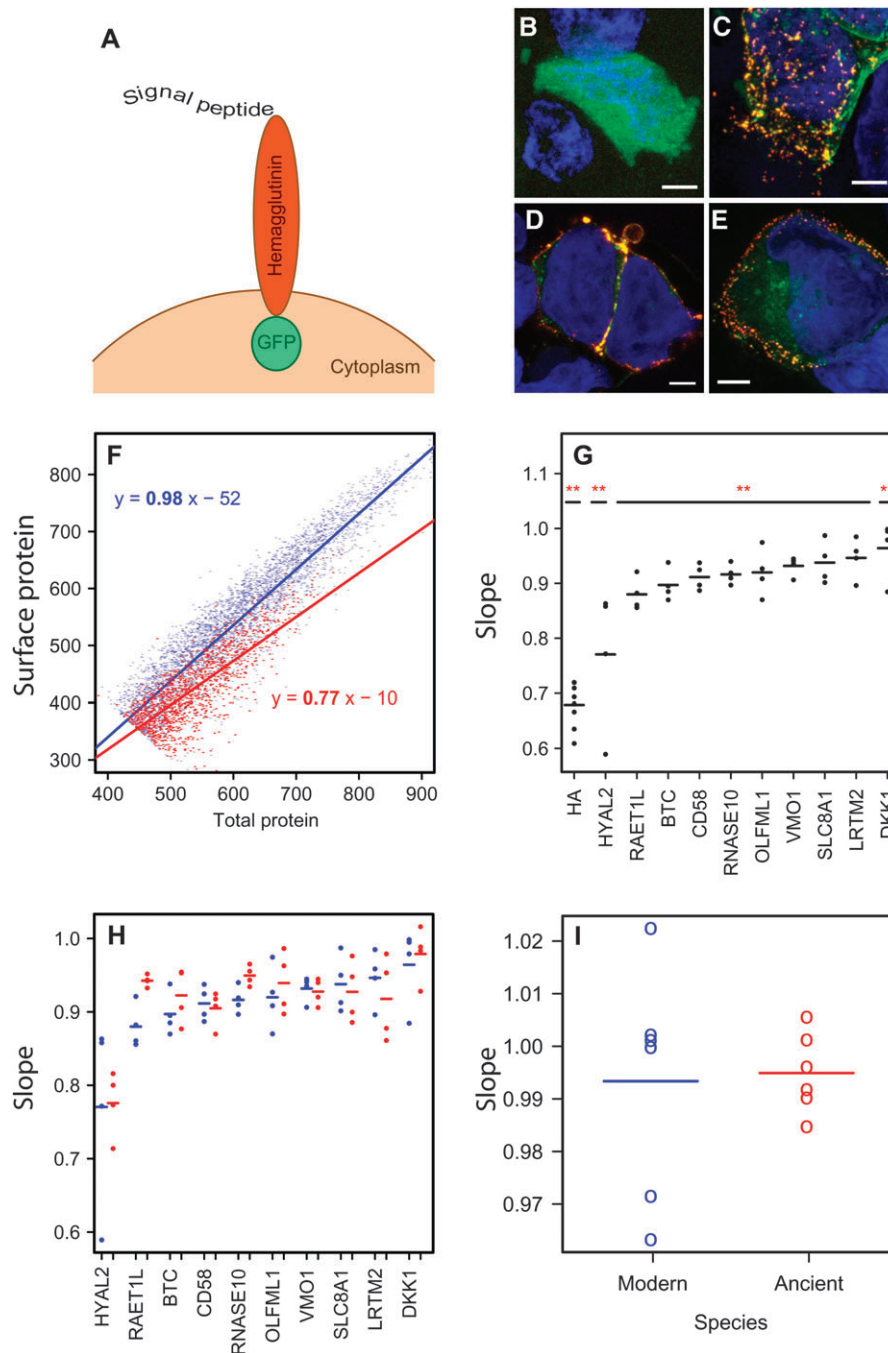
With the determination of a Neandertal genome sequence, it has become possible to identify amino acid substitutions that occurred on the human lineage either recently before or after its separation from the Neandertal lineage. The draft version of the Neandertal genome (Burbano et al. 2010; Green et al. 2010) has confirmed that the number of these substitutions is small, so that it will be possible in the future to functionally test all such amino acid substitutions as well as potentially other features in the genome that emerged in fully modern humans without

the reliance on model-dependent evolutionary tests. As a first application of this approach, we have investigated the evolution of signal peptides.

Signal peptides direct the cotranslational import of nascent proteins into the endoplasmic reticulum (ER). In the absence of other trafficking signals, membrane proteins carrying signal peptides are transported to the cell surface, whereas the signal peptides themselves are cleaved off after the successful insertion of the polypeptide into the membrane of the ER. Thus, with few exceptions, signal peptides serve only the function of directing proteins to the ER (Hegde and Bernstein 2006). A functional change in a signal peptide will therefore be reflected in a change of the efficiency with which mature membrane proteins reach the cell surface.

Changes in subcellular localization of proteins may be of evolutionary relevance (Marques et al. 2008), particularly because differences in the efficiency with which signal peptides direct the transport of membrane proteins to the cell surface can influence cellular survival and function (Hegde and Bernstein 2006; Kang et al. 2006), and several human diseases are due to single amino acid substitutions that abolish the function of signal peptides (Watzke et al. 1991).

Analysis of a data set where all nucleotide positions encoding amino acids that changed on the human evolutionary line were isolated and sequenced in a ~43,000-year-old Neandertal individual (Burbano et al. 2010) revealed ten cases where, in or near a signal peptide, the Neandertal carried an ancestral amino acid residue shared with chimpanzees, whereas the modern human reference genome encodes another derived variant (see [supplementary data, Supplementary Material](#) online). When 50 humans from around the world were analyzed, the derived variants were found to be fixed in four of these cases, whereas in the



**FIG. 1** Evolutionary analyses of signal peptide function. (A) Schematic illustration of a reporter construct containing a signal peptide (cleaved off after entering the ER), the mature sequence of the transmembrane protein hemagglutinin, and a C-terminal GFP. An antibody binds to the extracellular portion of hemagglutinin. (B–E) Confocal images of HEK293 cells expressing different reporter constructs. A mutant lacking a signal peptide is localized intracellularly (B), whereas a protein with a functional signal peptide reaches the plasma membrane where it is localized to speckles (C). The signal peptide from DKK1 causes most of the protein to localize to the plasma membrane (D), whereas the signal peptide from HYAL2 causes the protein to remain partly within the cell (E). (B) and (C) are maximum projections of the entire cells, whereas (D) and (E) exclude the ventral and dorsal surfaces of the cells. Green: GFP fluorescence from total reporter protein, red: antibody staining from surface reporter protein, and blue: nuclear stain. Scale bars: 5  $\mu$ m. (F) Fluorescence-activated cell-sorting data from representative experiments on constructs containing modern human DKK1 (blue) and HYAL2 (red) signal peptides. The equations describe the relation between total reporter protein (x) and surface reporter protein (y); the slope (bold) is a measure of signal peptide efficiency. (G) Efficiency of all modern human signal peptides tested. Dots indicate the slope for each independently stained and measured cell sample (two independent transfections on two different days), and horizontal lines indicate the mean value for each construct. **\*\*** $P < 0.01$  compared with all other groups. (H) Efficiency of the modern human signal peptides shown in (G) (blue) and the corresponding ancestral signal peptides (red). (I) Efficiency of the modern and ancestral versions of the signal peptide derived from RAET1L retested.

remaining six cases, both the ancestral and the derived variants exist among current humans (see [supplementary data, Supplementary Material](#) online). Each such modern/ancestral signal peptide pair was expressed in HEK293 cells as an N-terminal fusion to a reporter protein carrying a C-terminal green fluorescent protein (GFP) tag ([fig. 1A](#)). Reporter protein was visualized at the cell surface by antibody binding to nonpermeabilized cells, whereas total reporter protein was visible as GFP fluorescence ([fig. 1B–E](#)). A negative control with a signal peptide deletion (Sekikawa and Lai 1983) showed an exclusively intracellular localization ([fig. 1B](#)), whereas, in the presence of a signal peptide, the reporter protein reached the cell surface ([fig. 1C–E](#)); speckles on the plasma membrane showed that expression at the plasma membrane was not saturated ([fig. 1C](#)).

In order to quantify reporter protein on the entire surface of cells in a high-throughput manner, live cells were stained with the antibody and analyzed by fluorescence-activated cell sorting. We observed that for each construct, a linear relationship between total reporter protein and surface reporter protein was observed over the entire range of cell-to-cell expression variation. The slope of this linear relation, which is a good predictor of cell surface protein expression at all expression levels (median Spearman's  $r = 0.91$ ), differed significantly among signal peptides derived from distinct modern human genes ([fig. 1F and G](#)). As the mature reporter protein is the same in all cases (with the exception of six N-terminal amino acids, see [supplementary data, Supplementary Material](#) online), such differences are very unlikely to be due to differential folding of the mature protein in the ER or differential transport to the cell surface but to the interaction of the signal peptides with the translocation machinery or the propensity of the immature protein to be cleaved by signal peptidase.

However, none of the reporter proteins attached to a derived signal peptide differed significantly from its ancestral counterpart when the whole data set was considered ([fig. 1H](#); see [supplemental data, Supplementary Material](#) online, for statistical analysis). In order to further exclude that any such difference exists, the modern–ancestral pair with the greatest apparent difference in efficiency was reanalyzed in a further experiment ([fig. 1I, supplementary fig. 1, Supplementary Material](#) online). This confirmed that no significant difference between them exists. Furthermore, the four substitutions that are fixed in modern humans do not, as a group, differ in efficiency from the substitutions that are polymorphic in modern humans (see [supplemental data, Supplementary Material](#) online).

The lack of differences between the modern and ancestral reporter proteins in the cell line derived from embryonic kidney tissue shows that the signal peptide mutations that occurred on the modern human lineage had no general functional effects. Obviously, we cannot exclude that functional effects might exist in some cell type. However, the draft version of the Neandertal genome (Green et al. 2010) does not suggest any general changes in the protein

translocation or folding environment on the modern human lineage that might affect signal peptide function.

Our results are therefore most compatible with the neutral theory of molecular evolution (Kimura 1968). This holds true even for the four amino acid substitutions that are fixed in humans today. Although extrapolation from a particular class of substitutions may be dangerous, this suggests that many or most amino acid substitutions that have become fixed recently in human evolution may be without functional consequence. Nevertheless, the Neandertal genome will make it possible to perform comprehensive testing of all such substitutions in the future, as functional tests become available, and identify the presumably small fraction that does influence function. This will allow not only the identification of substitutions of functional significance but eventually also the reconstruction of aspects of the physiology of human ancestors at the time of divergence from Neandertals some 300,000 to 500,000 years ago.

## Supplementary Material

Supplementary data and [supplementary fig. 1](#) are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

We are grateful to Hernán Burbano, Roland Schröder, Roger Mundry, and Richard E. Green for help with experiments and analyses; to Robert A. Lamb (Department of Biochemistry, Northwestern University, IL) for hemagglutinin cDNA; to Steve Wharton and John Skehel (National Institute for Medical Research, London, UK) for the HC83X antibody; to Gunnar von Heijne (Department of Biochemistry and Biophysics, Uppsala University, Sweden) for several helpful suggestions; to Andreas Lösche, Viola Döbel, and Kathrin Jäger for cell sorting analyses performed at the Interdisciplinary Centre for Clinical Research, Leipzig University; to Susanna Schubert, Ingo Schäfer, and Peter Seibel for help and use of the confocal microscope at the Cellular Therapy Group, Biotechnology and Biomedicine Center, Leipzig University; and to the Presidential Innovation Fund of the Max Planck Society for financial support.

## References

- Akey JM. 2009. Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res.* 19:711–722.
- Burbano HA, Hodges E, Green RE, et al. (20-coauthors). 2010. Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* 328:723–725.
- Green RE, Krause J, Briggs AW, et al. (56-coauthors). 2010. A draft sequence of the Neandertal genome. *Science* 328:710–722.
- Hegde RS, Bernstein HD. 2006. The surprising complexity of signal sequences. *Trends Biochem Sci.* 31:563–571.
- Kang SW, Rane NS, Kim SJ, Garrison JL, Taunton J, Hegde RS. 2006. Substrate-specific translocational attenuation during ER stress defines a pre-emptive quality control pathway. *Cell* 127:999–1013.
- Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* 217:624–626.

- Marques AC, Vinckenbosch N, Brawand D, Kaessmann H. 2008. Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. *Genome Biol.* 9:R54.
- Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. *Science* 312:1614–1620.
- Sekikawa K, Lai CJ. 1983. Defects in functional expression of an influenza virus hemagglutinin lacking the signal peptide sequence. *Proc Natl Acad Sci U S A.* 80:3563–3567.
- Watzke HH, Wallmark A, Hamaguchi N, Giardina P, Stafford DW, High KA. 1991. Factor X Santo Domingo. Evidence that the severe clinical phenotype arises from a mutation blocking secretion. *J Clin Invest.* 88:1685–1689.