HaploGrep: A Fast and Reliable Algorithm for Automatic Classification of Mitochondrial DNA Haplogroups

Anita Kloss-Brandstätter,1∗† Dominic Pacher,2† Sebastian Schönherr,1,2† Hansi Weissensteiner,1† Robert Binna,2 Günter Specht,2 and Florian Kronenberg1

1Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria; 2Department of Database and Information Systems, Institute of Computer Science, University of Innsbruck, Innsbruck, Austria

Communicated by Pui-Yan Kwok
Received 5 July 2010; accepted revised manuscript 16 September 2010.
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/humu.21382

ABSTRACT: An ongoing source of controversy in mitochondrial DNA (mtDNA) research is based on the detection of numerous errors in mtDNA profiles that led to erroneous conclusions and false disease associations. Most of these controversies could be avoided if the samples' haplogroup status would be taken into consideration. Knowing the mtDNA haplogroup affiliation is a critical prerequisite for studying mechanisms of human evolution and discovering genes involved in complex diseases, and validating phylogenetic consistency using haplogroup classification is an important step in quality control. However, despite the availability of PhyloTree, a regularly updated classification tree of global mtDNA variation, the process of haplogroup classification is still time-consuming and error-prone, as researchers have to manually compare the polymorphisms found in a population sample to those summarized in PhyloTree, polymorphism by polymorphism, sample by sample. We present HaploGrep, a fast, reliable and straightforward algorithm implemented in a Web application to determine the haplogroup affiliation of all polymorphic positions of mtDNA profiles genotyped for the entire mtDNA or any part of it. HaploGrep uses the latest version of PhyloTree and offers an all-in-one solution for quality assessment of mtDNA profiles in clinical genetics, population genetics and forensics. HaploGrep can be accessed freely at http://haplogrepuibk.ac.at.


KEY WORDS: mitochondrial DNA; haplogroup; PhyloTree; quality assurance

Introduction

Human mitochondrial DNA (mtDNA) is routinely analyzed in various disciplines, such as medical genetics (searching for pathogenic mutations) [Durham et al., 2007; Elliott et al., 2008], population genetics (studying evolutionary patterns in human populations) [Kivisild et al., 2004; Pereira et al., 2009; Quintana-Murci et al., 2010], and forensic genetics (targeting degraded remains) [Biesecker et al., 2005; Ivanov et al., 1996]. The strict maternal inheritance of mtDNA results in a natural grouping of sequence haplotypes into monophyletic clusters, referred to as haplogroups. The members of a haplogroup carry a specific sequence motif as a consequence of sharing a common ancestor. All human mtDNA haplotypes can be traced back to a common matrilineal ancestor [Behar et al., 2008; Macaulay et al., 2005], and by assuming a sequential accumulation of mutations along maternally inherited lineages, all mtDNA haplogroups can be represented in a single phylogenetic tree. Currently, the most comprehensive overview of the human mtDNA tree is reflected in PhyloTree (http://www.phylotree.org), a regularly updated classification tree of global mtDNA variation considering all currently available whole mtDNA genome sequences [van Oven, 2010; van Oven and Kayser, 2008]. For population genetic inferences, the identification of haplogroup affiliation of members of a population sample is a crucial prerequisite for further investigations and conclusions on the demographic impact on human migrations. For clinical genetics, knowledge of the mitochondrial haplogroup can aid in finding disease-associated mutations, especially in case-control study settings in order to control for confounding due to mitochondrial population stratification [Biffi et al., 2010], and can help to distinguish them from either phylogenetically informative sites or sequencing artifacts.

It has been found that sample mix-up, contamination, biochemical problems, and transcription errors can lead to “phantom mutations” [Brandstätter et al., 2005], which in turn resulted in false associations and erroneous conclusions [Bandelt et al., 2004, 2007a]. As a consequence, guidelines concerning mtDNA genotyping were issued [Salas et al., 2005a]. These especially recommend double evaluation of mtDNA sequences, checking for phylogenetic consistency using haplogroup classification and the construction of quasi-median networks for the detection of phantom mutations [Bandelt et al., 2009; Brandstätter et al., 2006a].

However, despite the online availability of PhyloTree, a researcher who wants to determine the haplogroup affiliation of certain samples still has to manually compare the polymorphisms found in a population sample to those summarized in PhyloTree, polymorphism by polymorphism, sample by sample. This process is in most cases time-consuming, long-winded, and error-prone. A software tool for inferring the haplogroup of a (set of) sample(s) in an easy, straightforward and automated way would therefore be of immense practical importance.
We therefore developed HaploGrep, a Web application designed to determine the most probable haplogroup(s) of mtDNA haplotypes from the entire mitochondrial genome or any part of it. Using this application will significantly reduce the error rate and the amount of time needed for the preparation of those data for downstream quasi-median network analyses. Ultimately, HaploGrep yields an all-in-one solution for quality assessment of mtDNA profiles in clinical genetics, population genetics, and forensics.

Methods

Incorporation of Phylotree

Phylotree [van Oven and Kayser, 2008] is a clearly structured phylogenetic representation of the human mtDNA tree, with the potential to be re-coded into a computer-processable format like XML (Extensible Markup Language; http://www.w3.org/XML). Phylotree is stored the binary XLS format and provided for users as a HTML file. We designed a three-step model to generate the XML-version of Phylotree: (1) In a preprocessing step the provided XLS file is checked by VBA scripts to assure data integrity, because Phylotree is maintained in a manual fashion; (2) haplogroups, corresponding profiles, accession numbers, as well as references are extracted from the Excel file and encoded into an XML file, whereby the tree structure of Phylotree is preserved. As a general rule, mtDNA haplotypes are reported as a list of nucleotide positions that deviates from the revised Cambridge Reference Sequence (rCRS) [Andrews et al., 1999]. The rCRS is set as the base for the XML tree and the haplogroups are appended recursively; (3) this XML file is used to determine the phylogenetic weight of each polymorphism (see below). This procedure is performed as a service by our group each time a new version of Phylotree is published, which on average occurs every 6 months, to ensure that HaploGrep works with the most up-to-date knowledge of the human mtDNA phylogeny.

Determination of Phylogenetic Weights

HaploGrep’s approach to determine the most likely haplogroup includes the calculation of a phylogenetic weight for every mutation. As suggested by Soares et al. [2009], a site-specific mutation rate is needed to determine correct and reliable haplogroups. In particular, this rate is necessary if two haplogroups differ by only one single mutation. The weight is computed by traversing our XML representation of Phylotree to obtain the absolute number of occurrences per mutation in the entire tree. These numbers are used for calculating the weights of mutations needed by the HaploGrep algorithm by executing the following steps: (1) the mutations are sorted descendingly by their number of occurrences in the tree (e.g., m.152T>C—69; m.16189T>C—69; m.16519T>C: 107—67 = 40, etc.). As a final result, hotspot mutations with a high number of occurrences in the tree have a low phylogenetic weight, while stable mutations with only a few occurrences in the tree have a high phylogenetic weight. Super hot spots were not taken into consideration for the HaploGrep algorithm by assigning weight 0: m.523_524delAC, m.16182A>C, m.16183A>C, m.16519T>C, and C-insertions as m.309, m.315, and m.16193. Supp. Table S1 summarizes the occurrences of mutations in Phylotree and their corresponding weights in HaploGrep.

Main Algorithm

HaploGrep calculates a ranked list of relevant haplogroups for each given sample. To create this list, HaploGrep goes through all polymorphisms within a given sequence range from the human mtDNA haplogroup tree starting at the rCRS. The sequence range can include the entire mtDNA genome or any part of it, even single SNPs, in any combination of sequenced fragments and genotyped SNPs (e.g., 16024–16365; 70–340; 2706; 7028; 14766). For every haplogroup in Phylotree the adequate rank $r_{hg}$ is then calculated according to:

$$r_{hg} = \frac{1}{2} \times \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} w_i} + \frac{1}{2} \times \frac{\sum_{j=1}^{m} w_j}{\sum_{j=1}^{m} w_j},$$

where $w_i$ is the phylogenetic weight of the $i$th polymorphism, $k$ is the number of polymorphisms that were observed in the sample from the currently tested haplogroup, $m$ is the total number of polymorphisms expected in the analyzed range of the currently tested haplogroup, and $n$ is the total number of polymorphisms in the sample.

The formula shows that the rank calculation consists of two main components. The first part calculates the weighted ratio of haplogroup-associated polymorphisms that were found in the sample to the total number of haplogroup-associated polymorphisms in the currently tested haplogroup within the analyzed range. The second part reflects the ratio of polymorphisms in the test sample that are associated with the haplogroup under investigation to the total number of polymorphisms found in the test sample. This second component assures that a result will be ranked higher if it uses as many polymorphisms in the sample as possible. The two components are weighted equally and form the final overall rank $r_{hg}$. The resulting list is sorted according to these ranks to present the user the best results first.

For demonstration, we want to give an example: assume that a sample was sequenced for the entire control region (nucleotide positions 16024–16569; 1–576) and genotyped for 45 coding region SNPs as described by Brandstätter et al. [2006b]. It showed the following profile (with corresponding weights): m.263G>T, m.309_310insC (0), m.315_316insC (0), m.5700A>G (106), m.951G>A (103), m.8860A>G (106), m.13326A>G (106), m.16354C>T (101), and m.16351T>C (0). The polymorphisms m.309_310insC, m.315_316insC, and m.16359T>C were hot spots and therefore not taken into consideration (weighted with 0). For every haplogroup the rank was calculated according to HaploGrep’s algorithm: haplogroup H2a1 was rated best because all “diagnostic” polymorphisms for the haplogroup were found in the sample $[(105+106+103+106+106+101)/(105+106+103+106+101+101) = 1]$ and none of the sample’s polymorphisms remained unexplained or “private” $[(105+106+103+106+101)/(105+106+103+106+106+101+0+0+0) = 1]$. Therefore, the overall rank for H2a1 was $r_{H2a1} = 0.5\times1.0 + 0.5\times1.0 = 1.0$. For example, the third best hit was haplogroup H2a2, where the value for the first part of equation was still 1.0 but because two polymorphisms (m.951G>A, m.16354C>T) remained unexplained in the sample’s profile, the second part of the equation was $[(105+106+106+101)/(105+106+103+106+106+101+0+0+0) = 0.6746$, thus leading to a final rank of $r_{H2a2} = 0.5\times1.0 + 0.5\times0.6746 = 0.837$. [2]
Architecture

HaploGrep was designed as a Web application based on a client-server architecture. This allows the separation of the core application from the user interface and assures that the user always works with the latest version of HaploGrep.

The core of HaploGrep is implemented as a java library. REST was used for the client-server communication, implemented within the restlet framework [Fielding, 2000], thus putting all computational tasks on server side. The results of these tasks are provided as REST-resources using JSON (http://www.json.org) and XML as standard exchange formats. A server side session object is created for every new request and identified through an encrypted session key (please see “Security management”); all data are stored in an in-memory hash table, thus increasing the speed of HaploGrep significantly, because all operations are exclusively executed in main memory. Furthermore, this architecture utilizes current multicore architectures and allows the handling of concurrent requests efficiently.

The Client is implemented in JavaScript as HTML5-based rich internet application, communicating through asynchronous HTTP requests (AJAX) with the REST-resources. Using JSON as a transport medium within the JavaScript client allows a lightweight architecture and reduces the overhead. The user interface was implemented with the EXT-JS JavaScript framework (http://www.extjs.com). The JavaScript InfoVis Toolkit (http://thejit.org) was used for the visualization of Phylotree.

Workflow and Graphical User Interface

The import file for HaploGrep is a simple text file (with the file extension *.hsd) consisting of several columns: the first column contains the sample identifier (ID), the second column specifies the targeted genetic region (any sequenced range and any genotyped SNP, in any combination), the third column can include the assumed haplogroup status (or left blank if not specified), followed by the polymorphisms separated by tab stops (an example can be downloaded from HaploGrep’s homepage). This format can be created with spreadsheet applications such as Microsoft Excel or Open Office Calc and saved as a tab delimited text file. Alternatively, this format is generated automatically with the freely available epidemiological data management system eCOMPAGT [Schoenherr et al., 2009; Weissensteiner et al., 2010]. After importing this file into HaploGrep, the content is displayed in the upper part of the Web interface (Fig. 1) and the classification of haplogroups can be started; all samples are processed in batch mode. HaploGrep presents the top 10 results for every sample in the lower part of the Web interface; the top result is further used as recommended haplogroup. For each result, the computed final overall rank (see above) and the expected and unused polymorphisms are listed and can be surveyed in detail. The user can select any other haplogroup from the list of potential haplogroups as haplogroup status of the sample under investigation. Depending on the rank of the highest scoring haplogroup, the sample is highlighted in different colors.

Figure 1. HaploGrep’s Web interface. Notes: (1) Toolbar of HaploGrep; (2) data and associated haplogroups are highlighted in different colors depending on the quality of the assignment; (3) 10 top hits for every sample are displayed with details on diagnostic and remaining polymorphisms. The remaining polymorphisms are further categorized into “hotspot,” “local private mutation,” “global private mutation,” and “polymorphism out of range;” (4) location of the sample within the human phylogenetic tree; this visualization enables a view on polymorphisms that could additionally be analyzed for further refinement of the haplogroup status.
colors according to a traffic light scheme: samples with a suggested haplogroup that is ranked with a score above 0.9 are highlighted in green; scores between 0.8 and 0.9 are highlighted in yellow, and scores below 0.8 are emphasized in red color. In case a haplogroup was already defined in the input file, HaploGrep’s recommendation is added in brackets and the user needs to accept the changes manually. This conservative approach avoids an unreflected acceptance of results, as the user sometimes disposes of extra information on the samples’ ancestry. To improve the functionality of HaploGrep, an interactive visualization of Phylotree was created using the JavaScript InfoVis Toolkit. When selecting a haplogroup from the list of results, the entire phylogenetic branch is built and the polymorphisms are highlighted in different colors depending on whether they were typed in the sample or not and whether they have been found in the sample or not and even back-mutations are considered. The three tables in the lower left corner (see Fig. 1) show the results ranked by their weights, list the expected polymorphisms for the currently selected haplogroup, and indicate the status of the remaining polymorphisms that are not associated with the haplogroup. In the latter case we distinguish between “hotspot,” “local private mutation,” “global private mutation,” and “polymorphism out of range.” We define a polymorphism as “local private mutation,” when it is not associated with the current haplogroup, but has been observed in Phylotree at least once. On the contrary, a “global private mutation” refers to a polymorphism that is neither associated with the current haplogroup nor has it ever been seen in Phylotree. This information is very useful for data quality assurance.

After editing and selecting the appropriate haplogroup for each sample, the sample data can be saved as a tab-delimited text file or as an *.rdf file for the phylogenetic software Network.exe [Bandelt et al., 1999] (http://www.fluxus-engineering.com).

Security Management

HaploGrep neither saves nor stores user data or uploaded files. It runs on our server’s main memory where only a logging tool collects data for controlling for inappropriate software behavior and monitoring the status of HaploGrep. This supports the traceability of errors and improves the reliability of our application. HaploGrep’s session management is accomplished using HTTP’s cookie feature. Once a client has imported a file, the Web server assigns an encrypted session key (“source identifier”) to the client’s Web feature. Each time the client starts HaploGrep, the server checks if a session key is already available on client side. If the cookies were not deleted and the server session is still valid, the client has immediate access to his previously uploaded data. The session key is further encrypted by a cryptographic hash function (MD5); therefore, unauthorized clients cannot access foreign data. The session is deleted after 12 hr.

Results

Computation Time

The classification of haplogroups and the storage of several results per sample is a memory consuming task, which was evaluated on a multicore 64 bit Linux Server with 8 gigabyte RAM. HaploGrep scales very well in time and memory space, taking about 30 to 40 msec per sample to determine the haplogroup. If parallel requests are performed, HaploGrep makes use of the multi core architecture and distributes parallel requests to different cores, which leads to an even faster computation time (20 msec for each sample) compared to sequential requests. The memory usage scales in a linear way proportional to the number of imported samples (Fig. 2). With the current server configuration (8 GB of main memory) over 50,000 samples can be handled each day.

Validation Using Predetermined Data

A set of 100 previously published mtDNA profiles from the entire control region (16024–16569) (40 Europeans [Alshamali et al., 2008], 30 Asians [Alshamali et al., 2008; Brandstätter et al., 2007a; Zimmermann et al., 2009], and 30 Africans [Alshamali et al., 2008; Brandstätter et al., 2004]), which were randomly chosen from the supplementary material sections of the corresponding articles, were imported into HaploGrep (Supp. Table S2). Twenty-eight samples were assigned with the previously published haplogroup. Due to recent updates in Phylotree and the use of weights for each polymorphism, 59 samples were relabeled with a more precise haplogroup, and 13 samples were assigned a new haplogroup. However, it should be pointed out that at the time of publication of the selected test samples, the assigned haplogroup status reflected the state of knowledge at that time.

Finally, we compiled a dataset of 120 sequences with varying ranges (control region only [Brandstätter et al., 2004, 2006a, 2007a,b; Irwin et al., 2010; Lee et al., 2006], entire mitochondrial genome [Achilli et al., 2008; Behar et al., 2006, 2008; Pala et al., 2009; Palanichamy et al., 2004], and control region plus a variety of selected SNPs [Brandstätter et al., 2008; Zimmermann et al., 2009]) that is already available on HaploGrep’ Web application and that can be used as test dataset for understanding the functionality of HaploGrep without the need to upload own data (Supp. Table S3).

Comparison with Alternative Software Solutions

Besides HaploGrep, other software solutions exist to determine haplogroups: mtDNAManager provides a Web-based haplogroup classification for single and multiple mtDNA samples. However, this solution allows only the classification of haplogroups from the...
mtDNA control region (16024–16569; 1–576) [Lee et al., 2008]. HaploGrep differs from mtDNAmanager in several points: unlike mtDNAmanager the complete range of the mitochondrial genome is analyzed and the classification process is based on PhyloTree. This makes the classification far more reliable for users and provides a standard classification scheme. HaploGrep's results can be further verified either by analyzing the detailed results for every rank or by traversing the provided tree structure that makes the classification process fully transparent. A further important advantage is the increased accuracy and speed of HaploGrep compared to mtDNAmanager (Fig. 3). The test set described above was imported into HaploGrep and mtDNAmanager: only 43 samples were labeled identical by both systems; HaploGrep reassigned 46 samples with a more accurate haplogroup than mtDNAmanager (due to the use of PhyloTree); 11 samples were assigned an inappropriate haplogroup (given the genotyped range) by mtDNAmanager; 11 samples were reassigned 46 samples with a more accurate haplogroup than mtDNAmanager. HaploGrep is capable of determining about 20 samples per second and is therefore up to 28 times faster than mtDNAmanager.

MitoVariome is another database application that provides limited query functionalities to determine haplogroups using mutation motifs [Lee et al., 2009]. Unfortunately, it is only possible to classify one sample at a time and the classification process takes about 30 sec for a single sample. In addition, only the control region is examined and only the backbone of the human phylogenetic tree is used for guiding the classification process, which leads to a coarse assignment to only major haplogroup clusters.

**HaploGrep as a Tool for Quality Assurance**

The discovery of a wide range of errors in mtDNA disease studies [Bandelt et al., 2002, 2004, 2005, 2007a,b; Bandelt and Salas, 2009; Brandstätter et al., 2005; Kong et al., 2008; Salas et al., 2005b; Yao et al., 2003] has led to formulation of guidelines, which especially recommend checking for phylogenetic consistency using haplogroup classification and the construction of quasi-median networks for the detection of phantom mutations [Bandelt et al., 2001, 2009; Salas et al., 2005a]. A detailed instruction on how this can be undertaken is described in Bandelt et al. [2009]. Briefly, the first step is to allocate mtDNA profiles to major haplogroups; the second step is to ascertain whether all haplogroup-specific mutations in the new profile were actually observed, following the evolutionary pathway connecting the targeted sequence with the rCRS. HaploGrep can help to achieve this task especially through the visualization of the phylogenetic position of the sample in question. When, for example, the data of Uusimaa et al. [2004], which were discussed to be problematic in Bandelt et al. [2009], were imported into HaploGrep, it could be observed immediately that only a small part of samples had a score above 0.9 (Fig. 4). This would have been highly alarming, as the West Eurasian phylogeny is well known [Soares et al., 2010; Torroni et al., 2001] and especially entire mtDNA genomes should be easily located on a terminal node of the human mtDNA tree. A "phylogenetic bookkeeping of mutations" [Bandelt et al., 2009], that is, a systematic mutation by mutation comparison along known evolutionary pathways, as seen in Figure 4, would have helped to identify missed mutations and to check for phantom mutations on the list of "remaining" mutations, which were marked as "global private mutation."

**Discussion**

Accurate haplogroup assignment of mitochondrial haplotypes is an important but time-consuming task in human evolution studies and clinical genetics. Especially in the light of recent discoveries of a wide range of errors in mtDNA profiling, which resulted in false associations and erroneous conclusions [Bandelt and Salas, 2009; Bandelt et al., 2007b], it is necessary to check for phylogenetic consistency using haplogroup classification and to construct networks for the detection of phantom mutations [Bandelt and Dür, 2007]. However, the conversion of annotated mtDNA population data to Network.exe format [Bandelt et al., 1999] is a difficult and error-prone task. An automatic conversion can either be accomplished with HaploGrep or the recently described mtDNA data management software eCOMPAGT [Weissensteiner et al., 2010]. In addition, although PhyloTree provides a regularly updated and comprehensive phylogeny of global human mtDNA variation [van Oven, 2010; van Oven and Kayser, 2008], it is still necessary to have an insight into the mtDNA phylogeny to make use of this comprehensive information. Furthermore, a manual search for the presence or absence of diagnostic polymorphisms is a long-winded, complicated, and sometimes confusing task.

Studies of human populations worldwide are still far from being complete, and as a consequence, it is anticipated that new haplogroups will be discovered and that the already known haplogroups will be resolved further and characterized more precisely. These developments may lead to more precise results when performing haplogroup assignments of mtDNA profiles. However, haplogroup assignment will also become a more complicated and time-consuming process as more and more haplogroups need to be taken into consideration and this circumstance calls for automation. Additionally, with the increasing knowledge of variation in mitochondrial haplogroups, the need to reanalyze already typed samples will arise naturally. This task can only be achieved with reasonable time exposure when applying IT-based solutions. HaploGrep offers a fast, straightforward and reliable solution to this problem. The standardized and straightforward procedure of haplogroup assignment will also become a more complicated and time-consuming process as more and more haplogroups need to be taken into consideration and this circumstance calls for automation. Additionally, with the increasing knowledge of variation in mitochondrial haplogroups, the need to reanalyze already typed samples will arise naturally. This task can only be achieved with reasonable time exposure when applying IT-based solutions. HaploGrep offers a fast, straightforward and reliable solution to this problem. The standardized and straightforward procedure of haplogroup...
assignment with HaploGrep has the potential to overcome the issue of an intuitive assignment of samples to haplogroups. With each update of the classification tree Phylotree, the haplogroup classification procedure can quickly be repeated according to the actual state of known variation in the human mitochondrial genome.

HaploGrep is a freely available software solution with several strengths: (1) HaploGrep is implemented as Web application, and therefore no further installation procedures are required; (2) HaploGrep’s recommendations are based on Phylotree, a periodically updated classification tree estimated from data worldwide; (3) any given range of the mitochondrial genome can be used for haplogroup classification, which is based on the phylogenetic stability of mtDNA polymorphisms; (4) for every input sample the top 10 results and the phylogenetic position of the respective haplogroups are displayed, thus providing a detailed explanation of how and why a haplogroup was ranked best; (5) all results are visualized in a tree structure, showing the current position in Phylotree with possible hints for further refinement of the actual haplogroup status; (6) the classified data can be exported as a tab delimited text file or as a .rdf file for Network.exe. HaploGrep has limitations as well: its recommendations depend on the data accuracy of Phylotree, but because Phylotree is updated at least every 6 months, it is the most accurate presentation of haplogroups and therefore an excellent data source for HaploGrep. In addition, the accuracy of HaploGrep’s haplogroup classification depends on the analyzed sample range, clearly yielding more accurate results when using larger sequence ranges.

In agreement with the ongoing discussion on error prevention, we provide a constructive solution to check for phylogenetic consistency as recommended by Bandelt et al. [2009]. HaploGrep can be used without a login and imported samples are exclusively visible to the appropriate users due to HaploGrep’s session management. The export possibilities as a standard tab delimited file and as a .rdf file for the phylogenetic software Network.exe render HaploGrep currently the most comprehensive solution for human mtDNA haplogroup classification.

**Outlook**

Starting from now (Phylotree version 10), we will continuously update HaploGrep to work with the latest version of Phylotree. To strengthen the security of HaploGrep and to allow a secure data transfer, an encrypted channel will be available. This will be achieved by using HTTPS, a combination of HTTP with the SSL/TLS protocol, which guarantees protection from eavesdroppers.
Acknowledgments

Sebastian Schöntherr was supported by a scholarship from the University of Innsbruck (Doktoratsstipendium aus der Nachwuchsförderung, MIP10/2009/3). Hansi Weissensteiner was supported by a scholarship from the Autonomous Province of Bozen/Bolzano—South Tyrol. We thank Dr. Sarah J. Bourlat (Museum of Natural History, Stockholm, Sweden) for language revision of the article. The authors appreciate the constructive comments made by three anonymous reviewers, which considerably improved the application.

References


