Computer Note

Peditree: Pedigree Database Analysis and Visualization for Breeding and Science

R. VAN BERLOO AND R. C. B. HUTTEN

From the Laboratory of Plant Breeding, Wageningen University and Research Centre, P.O. Box 386, NL-6700 AJ Wageningen, The Netherlands.

Address correspondence to R. van Berloo at the address above, or e-mail: ralph.vanberloo@wur.nl.

At the Wageningen Laboratory of Plant Breeding, a software package has been developed to query a simple structured database with variety pedigree data. The package, called Peditree, creates a tree-shaped representation of pedigree information and has several visualization and lookup options. Estimates of inbreeding coefficient within a pedigree or coefficients of coancestry among pedigrees can be obtained. Furthermore trait data—if available—can be linked, displayed within the pedigree tree, and used to highlight pedigree entries that comply with set criteria.

Introduction

With the growing availability of molecular and biochemical data for plants, an old and valuable source of information tends to get a bit neglected. Plant variety pedigree information, records of the parental origin of genotypes that became cultivated or used themselves as parental lines for further breeding, provide a wealth of information on historic selection choices (e.g., Russel et al. 2000; Schut et al. 1997). Although the direct parents of a new variety are usually recorded, a truly extensive analysis of pedigree information requires a more structured view on the available information. Triggered by the availability of an elaborate set of potato pedigree data at the Wageningen Laboratory of Plant Breeding, we developed a software tool that allows more advanced pedigree analyses, using little more than the traditional parent–offspring relation tables as the data source. Such analyses can be used to give breeders an enhanced insight into their breeding materials, may provide data for better-founded selection choices, and can offer novel insights in the overall structure of the available germplasm currently and in the past.

The starting set of potato pedigree data (Swiezynski et al. 1997; supplemented by Hutten unpublished data) has been gathered from books, reports, records of breeding companies, and historic literature and was carefully screened for duplicates, errors, and other problematic data. The data set took the form of a spreadsheet file in which, for each potato variety, details about the parental origin, year of release, breeder, and other details were recorded. This facilitated a quick lookup of parental data given a genotype but did not allow a more complete description of the parental pedigree.

This limitation was somewhat resolved by transformation of the data set to a Web-accessible database (Van Berloo and Hutten 2000). The Web data set allowed lookup of parents of a potato variety and, by reporting the results as hyperlinks, the lookup of the parents of the parents, and so on. But manual navigation through the pedigree structure was required and still quite cumbersome. To eliminate these limitations, a stand-alone software package was developed.

Material and Methods

Peditree Software

The program was created in object Pascal using the Borland Delphi programming environment and runs under the Microsoft Windows environment. Data are provided in the form of a Microsoft Access database, containing a single table with the pedigree data. Within this table only three fields are required: genotype, parental origin of the genotype, and year of release, which can be left blank if unknown. Converting existing pedigree data to fit in this framework should be straightforward and not require more than the import of spreadsheet data and renaming of fields to fit the predefined names. The Peditree package uses SQL statements for data retrieval.

The main feature of the program is the lookup of a complete pedigree tree structure (hence the name Peditree). This is done recursively, which means that if a genotype A has parents B and C, the program will also look up the parents of B (say D and E) and also more complex parental

likelihoods specifically related to pedigree data that we needed to address:

• Duplicate use of names. This is best illustrated with an example: In one of our data sets we encountered a barley cultivar Opal (released 1998) that has deep down among its ancestors the cultivar Opal (released in 1926). Because lookup is text-based, an infinite loop could arise. By introducing a check for the year of release, this is avoided.

• Historic data are not always nicely recorded. We encountered parental data specifications in the form of A = (B × C) × E and also more complex parental
compositions. The lookup routine was modified to deal with these situations correctly and introduce an intermediate complex parent (B × C) when necessary.

- Some parental specifications needed special treatment, for instance, when “unknown” is specified as the parental origin, it is clear that we do not need to refer to a variety with name unknown, but that the data are missing. In this case and similar for descriptions like “mutant of,” “seedling of,” and “synonym of,” an appropriate action is taken, and pedigree building continues as well as possible.

Visualizations

The Peditree results of a single pedigree search can be displayed in a variety of ways. The default representation is an Explorer-like collapsible/expandable tree, allowing focus on specific parts of the pedigree while keeping other parts collapsed (see Figure 1 for an example).

The internal visualization was expanded with the external pedigree drawing routine Pedigraph (Garbe and Da 2003), which is able to visualize very complex pedigree relationships. Peditree exports data to a Pedigraph data file and launches Pedigraph, which then creates a conventional pedigree diagram. The Pedigraph diagrams are saved to disk but are also displayed automatically within Peditree. An example diagram is shown in Figure 2. For reference and further processing, the pedigree data can also be saved in a simple text file, using tab indents to indicate ancestral levels. An example of this output is displayed in Figure 3.

Calculations

In many pedigree structures common ancestors appear two or more times, often at different levels within the pedigree. This knowledge introduces the option to estimate an amount of common parentage or imbreeding, for the cultivar of interest. The method described by Falconer and Mackay (1996) for calculating the inbreeding coefficient (IBC) of a genotype with known pedigree was implemented, and the IBC is reported among the other pedigree results.

Comparisons among the pedigrees of genotypes and the retrieval of size estimates of a common origin (often called coefficient of coancestry) are included. These estimates can be gathered for one pair at a time, but a full diallel of a subset of varieties can also be analyzed in this way. Table 1 shows some results of this type of analysis.

Peditree allows users to perform a (calculation time-intensive) batch analysis of all genotypes within the data set. This results in a summary, listing for all cultivars the pedigree size, IBC, and so on, but also reporting for each cultivar the frequency of use within other pedigrees. This can be valuable knowledge for breeders and researchers, for instance in the identification of a representative set of germplasm.

Reverse Lookup

It is also possible to perform a reverse lookup, that is, to check in which other cultivar-pedigrees a cultivar of interest
appears as a progenitor. Again, a collapsible Explorer-like tree is used to display the lookup results. Because such an analysis can produce quite a large tree structure for popular progenitors, it is possible to limit the lookup to only a few levels, and continue deeper analysis step by step by simply clicking on nodes of cultivar names that need further evaluation. A routine to gather batchwise numerical data on the popularity as progenitor was recently added to the Peditree repertoire.

**Trait Data Linking**

If the user specifies a second database that contains trait data of (some of) the genotypes also present in the pedigree database, the Peditree software is able to display this trait data within the pedigree tree. Criteria for the trait data can be set, and the pedigree entries that comply with these criteria are shown highlighted. This option allows a visualization of transmission of, for example, increased disease resistance from one of the progenitors to the current cultivar. An example of this feature is shown in Figure 4.

**Future Development**

We plan to enhance the options available in Peditree in the future. One option might be marker visualization (possibly within the pedigree diagram). Earlier work on marker visualization (GGT; Van Berloo 1999) could provide a valuable starting place for this.

**Availability**

The Peditree software is available free of charge (after facultative registration) from the Web site of the Laboratory of Plant Breeding, Wageningen University, The Netherlands, www.dpw.wau.nl/pv/pub/peditree.

---

**Table 1.** Some results of the estimates of the relatedness due to common parentage (coefficients of coancestry) for a diallel between seven barley cultivars

<table>
<thead>
<tr>
<th></th>
<th>Aapo</th>
<th>Abelia</th>
<th>Acclaim</th>
<th>Agneta</th>
<th>Havila</th>
<th>Hellana</th>
<th>Horpoon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aapo</td>
<td>*</td>
<td>.125</td>
<td>.000</td>
<td>.002</td>
<td>.017</td>
<td>.004</td>
<td>.031</td>
</tr>
<tr>
<td>Abelia</td>
<td>*</td>
<td>.016</td>
<td>.002</td>
<td>.018</td>
<td>.004</td>
<td>.031</td>
<td></td>
</tr>
<tr>
<td>Acclaim</td>
<td>*</td>
<td>.000</td>
<td>.005</td>
<td>.003</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agneta</td>
<td>*</td>
<td>.015</td>
<td>.005</td>
<td>.004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Havila</td>
<td>*</td>
<td>.005</td>
<td>.021</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hellana</td>
<td></td>
<td></td>
<td></td>
<td>.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horpoon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

--

**Figure 2.** Pedigraph image of a simple pedigree structure. Pedigraph allows multiple occurrence of the same parent and will draw every parent only once but shows relations to multiple offspring if appropriate. Although Pedigraph images are created using external routines, they are constructed and displayed fully automatically within Peditree.
Figure 4. Example of Pedtree Explorer-style tree view with trait value highlighting active. A database containing resistance scores was coupled to Pedtree, and the highlight criterion was set to highlight varieties with a BYDV value > 2 (BYDV = barley yellow dwarf virus). For all cultivars with data on this trait, the value is indicated in the tree, but only the cultivars that match the selection criterion are highlighted using a background color (e.g., the third and fourth varieties from the top show two cultivars with data, but only Henni, with a BYDV > 2 gets highlighted).

References


Received September 30, 2004
Accepted January 5, 2005

Corresponding Editor: Reid Palmer