# PoMaMo—a comprehensive database for potato genome data

## Svenja Meyer*, Axel Nagel and Christiane Gebhardt[1]

RZPD German Resource Centre for Genome Research GmbH, Berlin, Germany and [1]Max-Planck-Institute for Plant Breeding Research, Cologne, Germany

## ABSTRACT

**A database for potato genome data (PoMaMo, Potato Maps and More) was established. The database contains molecular maps of all twelve potato chromosomes with about 1000 mapped elements, sequence data, putative gene functions, results from BLAST analysis, SNP and InDel information from different diploid and tetraploid potato genotypes, publication references, links to other public databases like GenBank (http://www.ncbi.nlm.nih.gov/) or SGN (Solanaceae Genomics Network, http://www.sgn.cornell.edu/), etc. Flexible search and data visualization interfaces enable easy access to the data via internet (https://gabi.rzpd.de/PoMaMo.html). The Java servlet tool YAMB (Yet Another Map Browser) was designed to interactively display chromosomal maps. Maps can be zoomed in and out, and detailed information about mapped elements can be obtained by clicking on an element of interest. The GreenCards interface allows a text-based data search by marker-, sequence- or genotype name, by sequence accession number, gene function, BLAST Hit or publication reference. The PoMaMo database is a comprehensive database for different potato genome data, and to date the only database containing SNP and InDel data from diploid and tetraploid potato genotypes.**

## INTRODUCTION

Genome analysis in potato (*Solanum tuberosum*) started with the construction of molecular linkage maps for the complement of twelve chromosomes, based on restriction fragment length polymorphism (RFLP) markers (1–7). At the Max-Planck-Institute for Plant Breeding Research, more than 1000 RFLP loci have been mapped in different mapping populations. Random potato genomic PstI restriction fragments, potato ESTs and cloned genes of known function were used as marker probes. RFLP mapping of tomato sequences in potato and potato sequences in tomato revealed high co-linearity between the genomes of these two closely related Solanaceae species and connected the different RFLP maps (1,3,6). Comparative mapping also identified conserved linkage blocks between the potato and *Arabidopsis* genomes (5). DNA sequence information has been collected for most of the markers employed in various mapping experiments. More recently, information on single nucleotide polymorphisms (SNPs) and insertion deletion polymorphisms (InDels) was generated at a number of loci, preferentially linked to genes controlling pathogen resistance (8). BAC (bacterial artificial chromosome) insertions have been anchored to the genetic map (8) and several BAC insertions have been fully sequenced (unpublished data). These genome data were the basis for localizing factors within the potato genome, which control agronomic characters relevant for cultivation and use of potato, such as disease resistance (9) and tuber quality traits (10,11).

To facilitate global access and use of these genome data, the PoMaMo (Potato Maps and More) database was constructed as a part of the GABI Primary Database (GabiPD), which is located at the RZPD German Resource Center for Genome Research (Berlin, Germany). GabiPD has been established as a central internet-based database within the German Plant Genome Project 'GABI' (Genomanalyse im biologischen System Pflanze), with the focus to collect and handle data from groups involved in GABI projects.

Here we report the basic structure and information content of this new database.

## DATABASE CONTENT

Twenty-four linkage maps, two for each potato chromosome, with altogether around 1000 mapped elements (RFLP loci or BAC clones), publication references, more than 2000 genomic and cDNA sequences from 30 different diploid and tetraploid potato genotypes, BLAST results, primer information for sequence amplification and more than 1600 SNP and InDel

*To whom correspondence should be addressed at Heubnerweg 6, D-14059 Berlin, Germany. Tel: +49 0 30 32639 261; Fax: +49 0 30 32639 111; Email: svenja@rzpd.de

positions have been integrated into a single relational database, which is accessible via internet.

The multi-level inheritance database schema allows the connection between different genomic data sets (for a detailed depiction of the PoMaMo database schema see http://gabi.rzpd.de/projects/Pomamo/PoMaMoDBSchema.shtml). This way, marker sequences from different potato genotypes, information on variable nucleotide positions, and details about marker chromosomal position were merged, for example. Mapped elements cannot only be retrieved by the name of the element, but also by sequence name, putative gene function, literature references, similarity to other annotated genes, or by GenBank/SGN sequence accession numbers for example.

## SEARCH AND VISUALIZATION INTERFACES

### GreenCards

The PoMaMo start page is accessible via https://gabi.rzpd.de/PoMaMo.html. The search and visualization interface Green-Cards can be called up directly from the PoMaMo entry page and enables the user to browse and search for a comprehensive set of potato genome data. GreenCards allows queries by genotype name (e.g. SR1), marker or sequence name (e.g. P1h3), keyword (i.e. function of annotated genes with a high similarity to the potato sequences, e.g. 'resistance') or sequence accession number (e.g. AJ487408). Wildcards can be used within the database searches. A GreenCards query using the keyword 'resistance', for example, provides the user with a list of search hits. All these hits can be selected for display on the web.

General information on the object such as genotypic information, details about the clone library, information on amplification or sequencing primers are shown at the top (Figure 1) and publication references are linked with PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi). For each element, which is mapped on potato chromosomes, the chromosome number and position in centimorgan is displayed. Since GreenCards is connected with the map visualization interface YAMB, it is possible to call-up the whole chromosomal map from a GreenCards page by a mouse-click (Figure 1). Sequence information is shown and links to GenBank entries via GenBank sequence accession numbers are realized or, in the case of tomato markers (TG markers), to SGN database. GreenCards displays also SNP and InDel positions as described in (8). The results from BLAST comparison against the non-redundant database from GenBank and the *Arabidopsis* protein database from MIPS (Munich Information Center for protein sequences, http://mips.gsf.de/proj/thal/db/) are accessible, i.e. the BLAST results are shown in summary or the whole BLAST report with a graphical overview of all hits can be called up by a mouse-click.

## YAMB—YET ANOTHER MAP BROWSER

The Java servlet tool YAMB was written to display genetic or physical maps of the potato chromosomes. The maps are directly accessible from the PoMaMo startpage or can be started from a GreenCards window as described above. The chromosomal positions of the mapped objects are read directly from the database and the maps are drawn dynamically. This way it is easy to add new elements and view the actual map at once. The chromosome maps constructed using populations BC916[2] and F1840 are shown in parallel, aligned and connected by marker loci mapped in both populations. The alignment shows that map distances between the same pairs of RFLP loci are variable between the different mapping populations. The F1840 map is with 1044 cM total length longer than the BC916[2] map with 922 cM. The differences observed can result from the different population size, the different parental genotypes and different sets of markers used for map construction. The maps can be zoomed out to view all mapped elements on a given chromosome at once or zoomed in for a more detailed view. The maps are interactive, that is a mouse-click on an element of interest opens up a GreenCards window, which displays all the information available for the element as described above. The literature referenced for each mapped element gives access to further genetic or biological data, for example whether a marker was used in mapping experiments for qualitative and quantitative resistance factors or was used for synteny studies between potato and tomato or between potato and *Arabidopsis*.

## DATA GENERATION

### Molecular linkage maps

Potato RFLP maps have been constructed based on diploid mapping populations, which originated from crossing non-inbred, heterozygous parental clones. The principles and algorithms for linkage group construction in this material have been described (12). The mapping populations BC916[2] (2–4) and F1840 (3,5,13) consisted of 67 backcross and 92 F1 individuals, respectively. In each population and for each chromosome, a maternal and paternal linkage group and linkage groups based on markers shared between both parents have been constructed, which are connected and aligned with each other via allelic bridges between maternal and paternal RFLP alleles (4,12). For simplicity of display in the PoMaMo database, the linkage groups for each chromosome of each population were merged, based on the arithmetic mean of the genetic distances between pairs of RFLP marker loci that had allelic bridges between the parental linkage groups (anchor loci). The other markers (informative for one parent only or shared among both parents) were ordered according to their genetic distance from flanking anchor loci. Marker order within closely linked groups of markers (<5 cM) should therefore be considered ambiguous.

### Sequence analysis, SNP and InDel detection

Marker plasmids with potato genomic (GP markers) or potato leaf cDNA (CP markers) insertions were subjected to single-run sequencing from both ends using vector-specific primers. Potato tuber ESTs (S and P markers originating from cultivar Saturna and Provita, respectively) were sequenced (single-run) from the 5′ end (5). Vector sequences were removed and overlapping forward and reverse sequences were assembled into single sequences. Custom sequencing was performed by the ADIS unit at the Max-Planck-Institute for Plant Breeding Research on ABI automated DNA sequencers (PE Biosystems,

## Clone: *P1h3*

**Genotype (Genotype)**

    **type:** wildtype

    **species:** Solanum tuberosum (potato)

    **cultivar:** Provita

**Library:**

    **lib name:** MPIZp2053

    **tissue:** tuber

    **cloning vector:** pBluescript SK- site_1: EcoRI; site_2: BamHI

**Mapping Info**

| map | species | chr. | mapped as | position | submitter |
|---|---|---|---|---|---|
| F1840 | Solanum tuberosum | 11 | *P1h3* | 24.001 cM | Frau Dr. Christiane Gebhardt |

**Reference:**

Gebhardt C, Walkemeier B, Henselewski H, Barakat A, Delseny M, Stüber K, 2003: The Plant Journal 34: 529-541 ID: 52

**Go to PubMed**

**Sequence:** *P1h3*

    **submitter of sequence:** Frau Dr. Christiane Gebhardt

        **length:** 667

        **GenBankAc:** AJ487408

```
P1h3
TCGTTTTCTC TCGATCGTTC GTTCGTTCAA GCTACTCTCA AGATGGGTGG AAGTGCGATA    60
AGTGATGATT GGGAATTCAC TACAACTGGA GCTCGGACAC TGGTTCTGGT TGGATGTACA   120
GGTAATGGTA AAAGTGCCAC AGGGAATAGC ATTGTTGGAA GAAAGGCATT CAAGTCAATG   180
TCTAGCTCAG CTGGTGTAAC AAGTACTTGT GAACTACAGA GGACTGTACT AGAAGATAAC   240
CAAATACTTG ATGTGATTGA TACACCAGGA TTGTTTGATT GTTCTTCTGA ACCTGAATTT   300
GTTGGAAACG AAATTGTTAA ATGCATCAAT ATGGCCAAGG ATGGGATTCA TGCTGTTCTT   360
GTAGTTTTAT CTGTGCGAAG TCGCTTTTCA AGAGAAGAAC ACGCAGCAGT CCAGAGTCTG   420
AGGGAGTTCT TTGGGAGCAA AATCAGTGAT TACATGATTT TGGTTTTCAC TGGTGGGGAT   480
GATCTTGAAG ATAATGATGA AACCCTGGAT GATTACTTAG GTCGTGACTG CCCCGAGCTT   540
TTAAAGAATA CCCTCAAGAT GTGTGGAAAT ANGCNANTGC TTTTCNATAA CAAGACTAAA   600
GAACCTTTGA ANAAACTGAA CAATTGAAAC AACTACTTTT CCTGGTTAAT GTTGTTGTAN   660
ANAATAN                                                             667
```

**Blast Searches:**

**MIPS arabi_all_proteins_v020204 on 15.03.04**

| E-val | simil. | overlap | description |
|---|---|---|---|
| 0.1e-47 | 50.7 | 195 | At1g33970 unknown protein |
| 0.9e-39 | 43.9 | 182 | At1g33930 AIG1-like protein |
| 0.4e-37 | 47.6 | 168 | At1g33960 AIG1 |
| 0.5e-35 | 40.9 | 171 | At1g33950 AIG1-like protein |
| 0.1e-34 | 42.8 | 189 | At4g09940 AIG1-like protein |

**All non-redundant GenBank CDS on 15.03.04**

| E-val | simil. | overlap | description |
|---|---|---|---|
| 0.1e-72 | 73.9 | 192 | gb|AAD09518.1| NTGP4 [Nicotiana tabacum] |
| 0.4e-46 | 50.7 | 195 | ref|NP_564431.1| avirulence-responsive protein, putative / avirulence induced gene protein, putative / AIG protein, putative [Arabidopsis thaliana] |
| 0.8e-46 | 50.7 | 193 | gb|AAN60299.1| unknown [Arabidopsis thaliana] |
| 0.8e-46 | 50.7 | 193 | pir||D86463 hypothetical protein F12G12.21 - Arabidopsis thaliana gb|AAG12538.1| Similar to AIG1 protein [Arabidopsis thaliana] gb|AAG12846.1| disease resistance protein AIG1; 5333-4002 [Arabidopsis thaliana] gb|AAM65167.1| AIG1-like protein, 5' partial [Arabidopsis thaliana] |
| 0.3e-37 | 43.9 | 182 | ref|NP_174655.1| avirulence-responsive family protein / avirulence induced gene (AIG1) family protein [Arabidopsis thaliana] pir||B86463 AIG1-like protein, 26931-23735 [imported] - Arabidopsis thaliana gb|AAG52203.1| AIG1-like protein; 26931-23735 [Arabidopsis thaliana] |

## Potato_Chromosome_XI

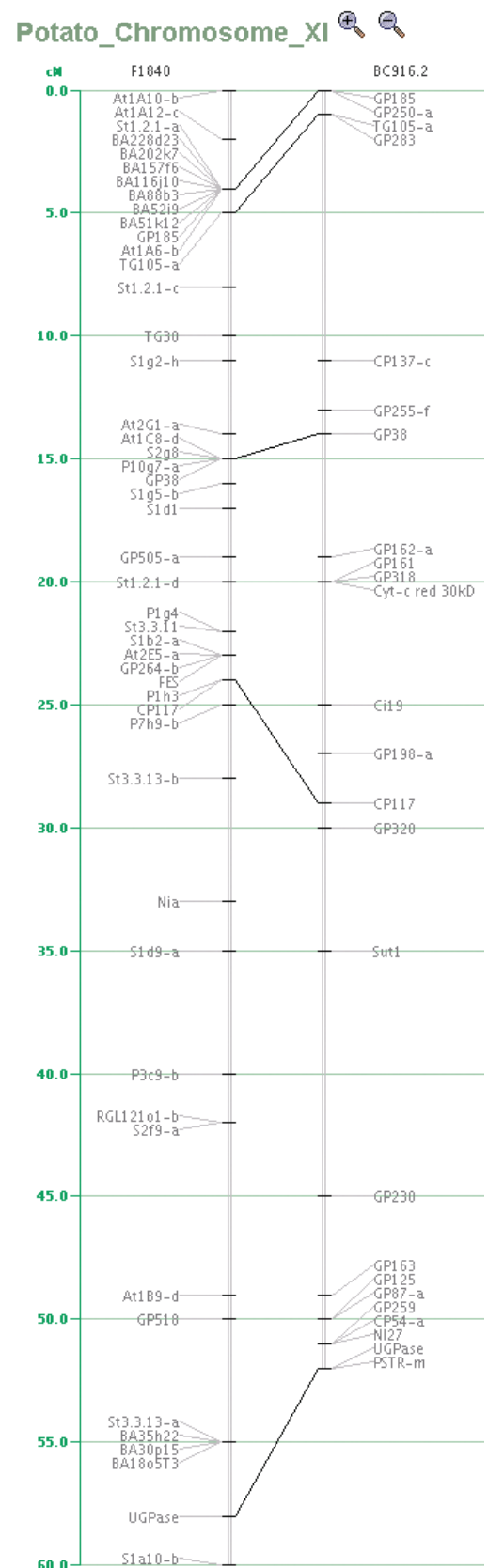| cM | F1840 | BC916.2 |
|---|---|---|
| 0.0 | At1A10-b, At1A12-c, St1.2.1-a, BA228d23, BA202k7, BA157f6, BA116j10, BA88b3 | GP185, GP250-a, TG105-a, GP283 |
| 5.0 | BA52i9, BA51k12, GP185, At1A6-b, TG105-a, St1.2.1-c | |
| 10.0 | TG30, S1g2-h | CP137-c |
| | At2G1-a, At1C8-d, S2g8 | GP255-f, GP38 |
| 15.0 | P10g7-a, GP38, S1g5-b, S1d1 | |
| 20.0 | GP505-a, St1.2.1-d | GP162-a, GP161, GP318, Cyt-c red 30kD |
| | P1g4, St3.3.11, S1b2-a, At2E5-a, GP264-b, FES | |
| 25.0 | P1h3, CP117, P7h9-b | Ci19 |
| | | GP198-a |
| | St3.3.13-b | CP117 |
| 30.0 | | GP320 |
| | Nia | |
| 35.0 | S1d9-a | Sut1 |
| 40.0 | P3c9-b | |
| | RGL121o1-b, S2f9-a | |
| 45.0 | | GP230 |
| | At1B9-d | GP163, GP125, GP87-a, GP259 |
| 50.0 | GP518 | CP54-a, NI27, UGPase, PSTR-m |
| 55.0 | St3.3.13-a, BA35h22, BA30p15, BA18o5T3 | |
| | UGPase | |
| 60.0 | S1a10-b | |

**Figure 1.** GreenCards and YAMB interfaces. Via GreenCards a data search can be started and detailed information on selected objects, like genotype and mapping information, publication references and sequence data as well as BLAST results are accessible. GreenCards is integrated with YAMB, which is a Java servlet tool for the display of mapping data.

Foster City, CA USA) using the dideoxy chain-termination sequencing method. SNP and InDel detection in diverse diploid and tetraploid potato genotypes was performed as described in (8). All sequences are also deposited in GenBank.

### Database implementation

The PoMaMo database is part of the GABI Primary Database. The website running on Apache Web Server (version 3.1)/ Tomcat (version 4.1) (http://www.apache.org/) is accessible via https://gabi.rzpd.de. The object-oriented database design was done using the data modelling application ER/Studio (version 5.5.2; Embarcadero, http://www.embarcadero.com/) and was implemented on a relational Oracle database (version 8.1.7; http://www.oracle.com) running on Tru64Unix (http://h30097.www3.hp.com/). Access to the database is realized via an object-oriented interface available in Perl and Java. The object-oriented interface is generated automatically from the database scheme. Data handling, i.e. insertions, updates and deletions, is done through this interface. Various Perl modules for processing special file formats (e.g. MS-EXCEL, GCG and CLUSTAL alignments, sequence data in FASTA, EMBL and other formats) were used. All potato sequence data are compared at regular intervals by BLASTX analysis (14) to the non-redundant (nr) protein database mirrored from GenBank and the annotated *Arabidopsis* protein database as available from MIPS. The BLAST results are integrated semi-automatically in the database. The Oracle InterMedia Text, which is a text-management and analysis extension to the Oracle database server, was utilized to allow database searches (with wildcards allowed) by keywords, gene functions, marker-, clone-, sequence- or cultivar-names, sequence accession numbers, publication references, etc. The web-accessible search and visualization interface GreenCards was written in Perl-cgi. Clone, SNP, primer and sequence information, BLAST results, literature references linked to PubMed, links to other databases, e.g. GenBank or SGN, are read directly from the database and are visualized. The Java servlet tool YAMB interactively displays linkage maps of the potato chromosomes. The maps were built dynamically considering information about the chromosomal positions of each element as read from the database. GreenCards and YAMB are integrated with each other.

### DISCUSSION AND CONCLUSION

The PoMaMo database harbours to date a comprehensive collection of genomic potato data, like sequences, SNP/InDel information and mapping data. It is the only database that contains information on SNP and InDels from diploid and tetraploid potato genotypes. The object-oriented structure of the database makes it possible to enlarge the schema with regard to new types of data, e.g. phenotypic data, and to merge phenotypic with genotypic data.

The GabiPD structure will also allow the quick integration of physical and function mapping data, gene expression and proteomics data, once these data become available for potato.

All data visualized via the web-accessible interfaces Green-Cards and YAMB were selected directly from the database; in this way, data updates and newly integrated data are accessible at once. Via the networking of YAMB and the text-based

search and visualization interface GreenCards, it is easy to switch from the map display to the GreenCards view, which provides detailed information on objects of interests or vice versa. Due to the modular body of GreenCards, the tool is extendable to upcoming data types, like phenotypic information. YAMB allows the interactive and parallel depiction of two or more maps, which are connected by identical markers. YAMB is, therefore, also a very comfortable tool to display synteny maps.

The sequence-tagged sites in the potato genome, as retrievable from the PoMaMo database, provide a resource for the mapping and marker-assisted selection of phenotypic traits in potato, tomato and other related species of the Solanaceae family. They can also be used for anchoring the potato genetic maps including function maps to physical maps of potato and other Solanaceae species and to whole-genome sequences of other plants, thereby connecting structural with functional genome analysis. The PoMaMo database can also function as one module in a global network of plant genome databases.

### AVAILABILITY

PoMaMo is accessible at the URL http://gabi.rzpd.de/PoMaMo.html. Inquiries related to the database should be directed to gabi@rzpd.de.

### ACKNOWLEDGEMENTS

### REFERENCES

1. Bonierbale,M., Plaisted,R.L. and Tanksley,S.D. (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics*, **120**, 1095–1103.
2. Gebhardt,C., Ritter,E., Debener,T., Schachtschabel,U., Walkemeier,B., Uhrig,H. and Salamini,F. (1989) RFLP analysis and linkage mapping in *Solanum tuberosum*. *Theor. Appl. Genet.*, **78**, 65–75.
3. Gebhardt,C., Ritter,E., Barone,A., Debener,T., Walkemeier,B., Schachtschabel,U., Kaufmann,H., Thompson,R.D., Bonierbale,M.W., Ganal,M.W. *et al.* (1991) RFLP maps of potato and their alignment with the homologous tomato genome. *Theor. Appl. Genet.*, **83**, 49–57.
4. Gebhardt,C., Ritter,E. and Salamini,F. (2001) RFLP map of the potato. In Phillips,R.L. and Vasil,I.K. (eds), *Advances in Cellular and Molecular Biology of Plants: DNA-based Markers in Plants*. Kluwer Academic Publishers, Dordrecht, Boston, London, Vol. VI, pp. 319–336.
5. Gebhardt,C., Walkemeier,B., Henselewski,H., Barakat,A., Delseny,M. and Stüber,K. (2003) Comparative mapping between potato (*Solanum tuberosum*) and *Arabidopsis thaliana* reveals structurally conserved domains and ancient duplications in the potato genome. *Plant J.*, **34**, 529–541.
6. Tanksley,S.D., Ganal,M.W., Prince,J.P., de Vicente,M.C., Bonierbale,M.W., Broun,P., Fulton,T.M., Giovannoni,J.J., Grandillo,S., Martin,G.B. *et al.* (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics*, **132**, 1141–1160.
7. Jacobs,J.M.E., Van Eck,H.J., Arens,P., Verkerk-Bakker,B., te Lintel Hekkert,B., Bastiaanssen,H.J.M., El Kharbotly,A., Pereira,A., Jacobsen,E. and Stiekema,W.J. (1995) A genetic map of potato

(*Solanum tuberosum*) integrating molecular markers, including transposons, and classical markers. *Theor. Appl. Genet.*, **91**, 289–300.

8. Rickert,A.M., Kim,J.H., Meyer,S., Nagel,A., Ballvora,A., Oefner,P.J. and Gebhardt,C. (2003) First-generation SNP/InDel markers tagging loci for pathogen resistance in the potato genome. *Plant Biotechnol. J.*, **1**, 399–410.

9. Gebhardt,C. and Valkonen,J.P.T. (2001) Organization of genes controlling disease resistance in the potato genome. *Annu. Rev. Phytopathol.*, **39**, 79–102.

10. Schäfer-Pregl,R., Ritter,E., Concilio,L., Hesselbach,J., Lovatti,L., Walkemeier,B., Thelen,H., Salamini,F. and Gebhardt,C. (1998) Analysis of quantitative trait loci (QTL) and quantitative trait alleles (QTA) for potato tuber yield and starch content. *Theor. Appl. Genet.*, **97**, 834–846.

11. Menendez,C.M., Ritter,E., Schäfer-Pregl,R., Walkemeier,B., Kalde,A., Salamini,F. and Gebhardt,C. (2002) Cold-sweetening in diploid potato. Mapping QTL and candidate genes. *Genetics*, **162**, 1423–1434.

12. Ritter,E., Gebhardt,C. and Salamini,F. (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics*, **125**, 645–654.

13. Leister,D., Ballvora,A., Salamini,F. and Gebhardt,C. (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nature Genet.*, **14**, 421–429.

14. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.