Mol Breeding (2013) 31:655–663
DOI 10.1007/s11032-012-9823-7

TGB: the tobacco genetics and breeding database

Hengchun Cao · Yujun Wang · Zhixin Xie · Lisha Huang · Houjuan Xu · Li Zhang · Ming Bian · Guangwei Sun · Shuaishuai Han · Long Yang

Received: 21 May 2012 / Accepted: 3 December 2012 / Published online: 22 December 2012

Abstract The germplasm of the genus *Nicotiana* contains more than 5,000 accessions and plays an important role in modern biological research. Tobacco can be used as a model system to develop methodologies for plant transformation and for investigating gene function. In order to develop the study of *Nicotiana*, a large quantity of data on germplasm, sequences, molecular markers and genetically modified tobacco was required for in-depth and systematic collation and research. It became necessary to establish a special database for tobacco genetics and breeding. The tobacco genetics and breeding (TGB, http://yancao.sdau.edu.cn/tgb) database was developed with the aim of bringing together tobacco genetics and breeding. The database has three main features: (1) a materials database with information on 1,472 *Nicotiana* germplasm accessions, as well as updated genomic and expressed sequence tag (EST) data available from the public database; (2) a molecular markers database containing a total of 12,388 potential intron polymorphisms 10,551 EST-simple sequence repeat (EST-SSR) and 66,297 genomic-SSR markers; and (3) an applications database with genetic maps and some genetically modified studies in tobacco. The TGB database also makes Basic Local Alignment Search Tool and primer designing tools publicly available. As far as can be ascertained, the TGB database is the first tobacco genetics and breeding database to be created, and all this comprehensive information will aid basic research into *Nicotiana* and other related plants. It will serve as an excellent resource for the online tobacco research community.

H. Cao · Y. Wang · H. Xu · L. Zhang · M. Bian · G. Sun · S. Han · L. Yang (✉)
Tobacco Lab, College of Plant Protection, Shandong Agricultural University, Daizong Road No. 61, Taian 271018, Shandong, China
e-mail: lyang@sdau.edu.cn

Y. Wang
e-mail: yancao@sdau.edu.cn

H. Xu
e-mail: xhjuan@sdau.edu.cn

L. Zhang
e-mail: lilizhang324@163.com

M. Bian
e-mail: bianmingbm@163.com

G. Sun
e-mail: sgw_dw@163.com

S. Han
e-mail: hanss0816@126.com

Z. Xie
College of Information Science and Engineering, Shandong Agricultural University, Daizong Road No. 61, Taian 271018, Shandong, China
e-mail: zxxie@sdau.edu.cn

L. Huang
College of Agronomy, Shandong Agricultural University, Daizong Road No. 61, Taian 271018, Shandong, China
e-mail: huang2002003@gmail.com

Keywords Tobacco · Germplasm · Molecular markers · Genetics · Breeding · Database

Abbreviations

TGB	Tobacco genetics and breeding
EST	Expressed sequence tag
BLAST	Basic Local Alignment Search Tool
PIP	Potential intron polymorphism
SSR	Simple sequence repeat
GM	Genetically modified
QTL	Quantitative trait loci
MAS	Marker-assisted selection

Introduction

Tobacco (*Nicotiana tabacum* L.) is a member of the agriculturally important Solanaceae and is grown as a commercial crop in many different countries. It is also widely used in plant breeding and genetics research (Rushton et al. 2008). As a model system in scientific research, it plays an important role in the study of phenotypic diversity, hybridization and ploidy manipulations, tissue culture, plant transformation and gene function investigation (Lewis 2011).

The genus *Nicotiana* was established by Linnaeus in 1753 and is a relatively large genus. For a long time, *Nicotiana* had been divided into three subgenera, 14 sections, and 66 species (Goodspeed and Thompson 1945, 1959), but with the development of taxonomy most scientists now tend to classify *Nicotiana* into 13 sections and 76 naturally occurring species (Lewis 2011; Knapp et al. 2004). Tobacco is a natural allotetraploid ($2n = 4x = 48$) formed through hybridization between two diploid ($2n = 24$) progenitors, *N. sylvestris* and *N. tomentosiformis* (Ganapathi et al. 2004; Leitch et al. 2008). Genetic resources provide the basic material for selection and improvement through breeding and genetic research. Some studies have shown that there was a narrow genetic background in *Nicotiana* species and a high genetic similarity between cultivated tobacco (Moon et al. 2008). Understanding and utilizing the available genetic diversity in cultivated and wild *Nicotiana* species are essential for the improvement of cultivated tobacco, continued genetic modification and other fundamental studies in plant biology (Lewis 2011).

The conserved germplasm has been characterized for important agronomic characters and germplasm seed samples have been distributed to researchers for utilization in crop improvement. Owing to the spread of modern cultivars, genetic variation is fast disappearing because the natural habitats of the wild relatives of the cultivated species are being destroyed (Upadhyaya et al. 2008). In order to maintain a high genetic diversity in tobacco, there has been a great deal of attention applied to collecting and maintaining *Nicotiana* germplasm all over the world (Day-Rubenstein and Heisey 2003; Moon et al. 2008, 2009a, b; Ravisankar et al. 2008). In the past 20 years, the major effort in breeding has changed from traditional phenotypic pedigree-based selection systems to molecular genetics, with emphasis on quantitative trait loci (QTL) identification and marker-assisted selection (MAS). DNA markers can be used to select optimal genotypes and are an excellent tool for selecting beneficial genetic traits that are difficult to measure, exhibit low heritability, and/or are expressed late in development (Ribaut and Hoisington 1998; Ruan et al. 2010; Wilde et al. 2007). To date, there has been a large number of molecular markers applied in tobacco research, such as RFLP (restriction fragment length polymorphism) (Martz et al. 1998), RAPD (random amplified polymorphic DNA) (Bai et al. 1995), AFLP (amplified fragment length polymorphism) (Bai et al. 1995; Ren and Timco 2001) and ISSR (inter-simple repeat sequence) (Denduangboripant et al. 2010).

In order to study tobacco in depth, the worldwide tobacco genome project was developed in several countries, such as the Tobacco Genome Initiative in America, which generated both the genomic (methyl-filtered and bacterial artificial chromosome) sequences as well as expressed sequence tags (ESTs) for *Nicotiana tabacum* (TGI, http://www.pngg.org/tgi/), and the ES-Tobacco project in Europe, which contains the EST sequences of four types of *Nicotiana* species (http://www.estobacco.info/). However, the huge size of the tobacco genome (3.5–5 Gb) makes the goal of sequencing the complete tobacco genome difficult. To date, TGI has increased the sequence information available on the open reading frames of the *N. tabacum* genome. Simultaneously, PlantGDB has made available a large amount of EST data for four *Nicotiana* species (http://www.plantgdb.org/).

With so much original and analyzed data, it became necessary to establish a special database for tobacco to

store and manage the biological information. At present, there are many plant resource databases on the Web, especially for single plant species, such as maizeGDB (http://www.maizeGDB.org) for *Zea mays* (Lawrence et al. 2004) and the Arabidopsis Information Resource (TAIR, http://www.arabidopsis.org) for *Arabidopsis thaliana* (Rhee et al. 2003). However, most tobacco databases have mainly concentrated on transcription factors (TOBFAC) (Rushton et al. 2008), gene expression (TobEA) (Edwards et al. 2010) or a germplasm database in Chinese (Zhang et al. 1990). There are also several comprehensive databases for Solanaceae, for example the Solanaceae Genomics Network (SGN, http://solgenomics.net/) and Solanaceae Genomics Resource (SGR, http://solanaceae.plantbiology.msu.edu/), both of which mainly focus on genomic analysis and concentrate on potatoes and tomatoes. With the aim of creating integrated data utilization and convenient access, an integrated and systematic database—the tobacco genetics and breeding database (TGB, http://yancao.sdau.edu.cn/tgb)—was developed. It is a comprehensive platform covering a broad spectrum of data, from genetics to breeding research data, which are needed for the higher model plant of *Nicotiana*. Currently, the TGB database contains a materials database (germplasm and genomes), a molecular markers database [potential intron polymorphism (PIP) and simple sequence repeat (SSR)], an applications database (genetic maps and genetically modified information) and some useful tools for tobacco research. The TGB database supports geneticists and breeders in their genetic studies and in their exploration of germplasm collections and is expected to benefit *Nicotiana* and wider related research.

Constructions and content

Architecture and implementation

The TGB database consists of some interrelated relational databases implemented in MySQL. The data handling and analysis parts of the database used the pipelines in Perl Script. The web interface was implemented in HTML running on an Apache web server. The TGB database has been set up on a World-Wide Web server allowing internet access with a web client.

Content

Three components make up the TGB: a materials database, a molecular markers database and an applications database (Fig. 1). Over the past 50 years, some advanced methodologies for collecting, maintaining, evaluating and documenting tobacco germplasms have been developed. Now, all data have been stored in the TGB database. In the materials database, there are 1,472 *Nicotiana* varieties (Table 1) covering nearly all *Nicotiana* species, with detailed information (e.g. economic characters, physiological and biochemical properties) on each. In order to include a wide spread of genomic research data, the current sequence data were downloaded from public databases (PlantGDB and TGI). Based on the germplasm and publicly available sequence data, the TGB database program developed the molecular markers database, which contains a large number of PIP, EST–SSR and genomic-SSR markers. It includes almost all the PIP and SSR markers available for tobacco study and could benefit both genetics and breeding research. Using the data in the materials and molecular markers databases, a genetic map was generated based on microsatellites (Bindler et al. 2011), and some important GM (genetically modified) studies with *Nicotiana* species have been summarized and collated in the applications database. Furthermore, some useful tools have been exploited for BLAST and for developing primers in this site. All of the data and tools can be downloaded freely from the TGB database.

Utility

The materials database in TGB

Germplasm

It has been recognized that germplasm has a basic role to play in the improvement of cultivated plants (Hawkes 1977). The information on the 1,472 *Nicotiana* varieties stored in the TGB database has been useful for tobacco breeding and genetic research. This information can be searched in the following ways (Fig. 2): (1) Type a keyword of the ID (e.g. SNT1) or the name (e.g. K324). (2) Select the type of tobacco (e.g. flue-cured tobacco). Then, click the button 'Submit' and the result with the following information

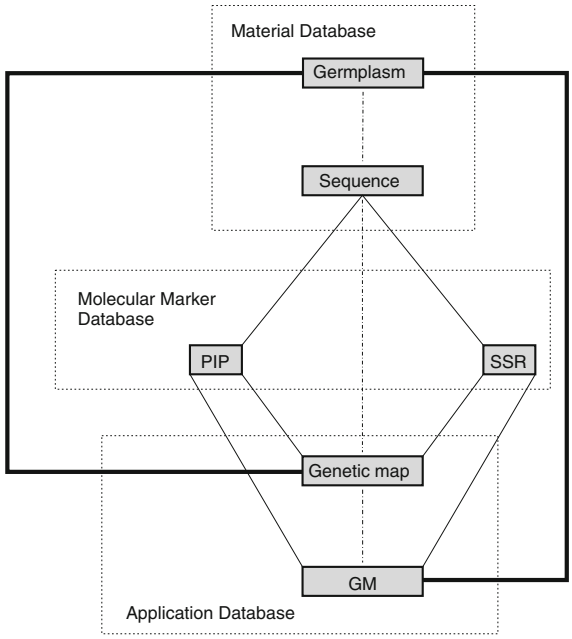


Fig. 1 TGB data relationship diagram. The Germplasm and Sequence databases are the foundation material databases for other TGB databases. These include the molecular markers database (PIP and SSR) and the Applications database (genetic maps and genetically modified information)

Table 1 TGB Germplasm database statistics

Type	Number
Flue-cured tobacco	933
Air-cured tobacco	2
Burley	98
Cigar	56
Wild	42
Rustica tobacco	31
Original tobacco	46
Sun-cured tobacco	172
Total	1,380

appears: ID, name and type of a tobacco. After clicking the name, users can obtain detailed information (Fig. 2), which includes a total of 48 relevant messages in a table as follows: (a) genetic background, e.g. family name, generic name and specific name; (b) economic characteristics, e.g. plant height, flower and leaf characteristics; (c) physiological and biochemical properties, e.g. protein, sugar and Schmuck's value, and (d) resistance characteristics: e.g. resistance to CMV, black-shank etc.

Fig. 2 The home page of the TGB and the search and development pages. This figure shows the home page **a** of the TGB database, the germplasm (**b**) search page and the SSR (**c**) and PIP (**d**) development pages. The home page **a** is the main entry page, providing quick access to resources through graphical menus. Every TGB database page consistently contains the same toolbar at the top with pull-down menus and links to download and tools pages. Links to other important resources are also provided. In particular, after a user runs a search for a germplasm, the result pages will show the detailed information for the chosen germplasm (**b**). Alternatively, on running a development for the SSR (**c**) or PIP (**d**) markers, the exact results will be available if the primers exist in the query sequences

Along with the tobacco germplasm experiments, the TGB database will update information in order to improve the quality of the germplasm data, such as pictures for each *Nicotiana* variety. In order to foster research and improved information exchange, the TGB database distributes germplasm to support study and educational objectives.

Sequence data

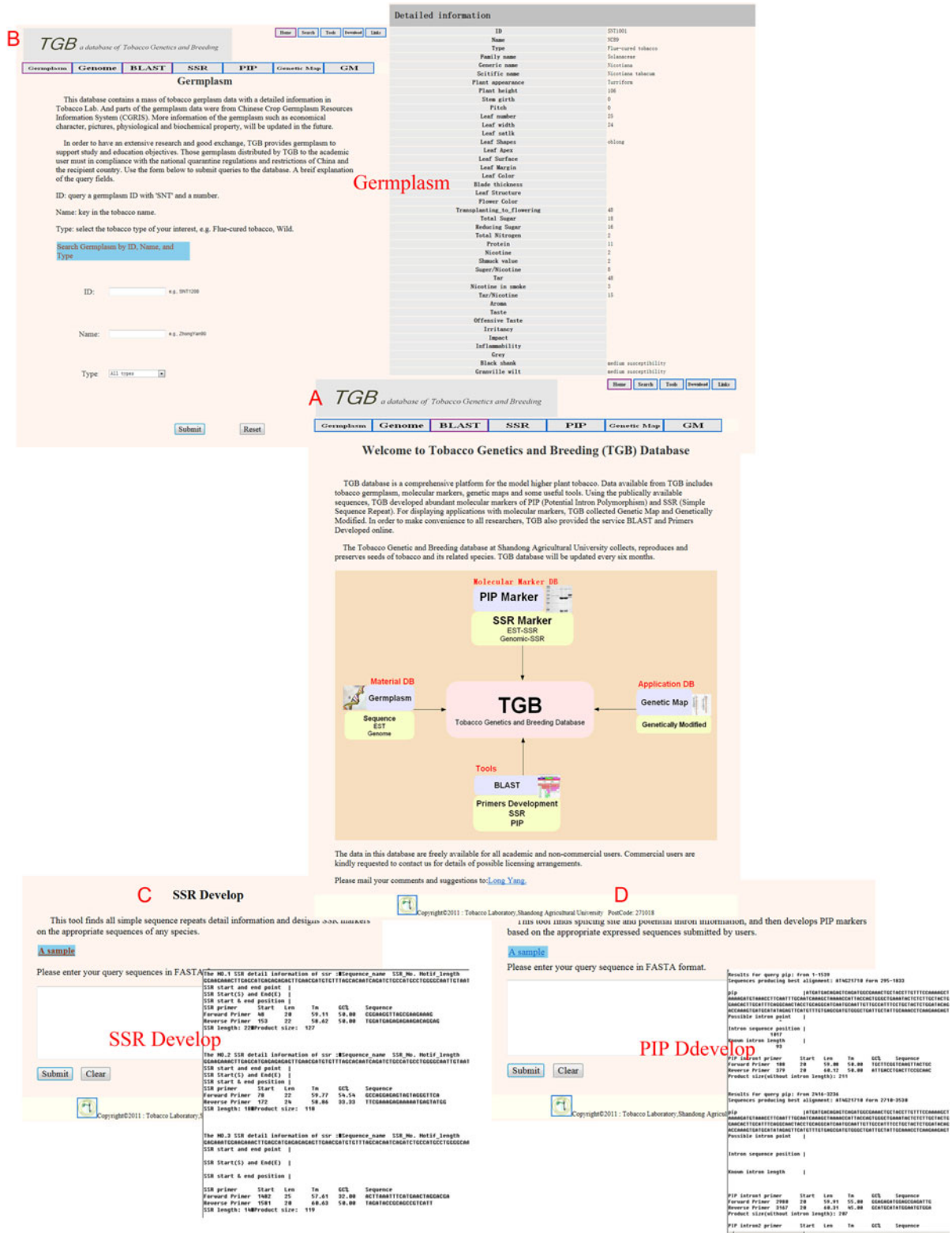
A total of 1,223,537 accessions of *N. tabacum* genomic sequences and 171,570 EST sequences of four *Nicotiana* species (*N. tabacum, N. sylvestris, N. benthamiana* and *N. langsdorffi* × *N. sanderae*) were downloaded from the TGI and PlantGDB databases, respectively. The sequence data were stored in two tobacco databases (coding sequences and genome sequences) on the BLAST page. Using those databases, scientists can finish the BLAST and design primers in their respective web pages. Furthermore, new sequence data will be regularly updated into the TGB database as new data is acquired from the public database.

The molecular markers database in TGB

A large number of PIP, EST-SSR and genomic-SSR markers based on known tobacco sequences have been developed and deposited in the TGB database.

PIP markers

Introns are non-coding sequences interspersed in genes. Subjected to less general selective pressure, introns are more variable than exons. It is not possible to define the tobacco intron structures at present as its


B
TGB a database of Tobacco Genetics and Breeding
Germplasm
Genome
BLAST
SSR
PIP
Genetic Map
GM
Germplasm
This database contains a mass of tobacco germplasm data with a detailed information in Tobacco Lab. And parts of the germplasm data were from Chinese Crop Germplasm Resources Information System (CGRIS). More information of the germplasm such as economical character, pictures, physiological and biochemical property, will be updated in the future.
In order to have an extensive research and good exchange, TGB provides germplasm to support study and education objectives. Those germplasm distributed by TGB to the academic user must in compliance with the national quarantine regulations and restrictions of China and the recipient country. Use the form below to submit queries to the database. A breif explanation of the query fields.
ID: query a germplasm ID with 'SNT' and a number.
Name: key in the tobacco name.
Type: select the tobacco type of your interest, e.g. Flue-cured tobacco, Wild.
Search Germplasm by ID, Name, and Type
ID:
Name:
Type
All types
Submit
Reset
Germplasm
Detailed information
A
TGB a database of Tobacco Genetics and Breeding
Welcome to Tobacco Genetics and Breeding (TGB) Database
TGB database is a comprehensive platform for the model higher plant tobacco. Data available from TGB includes tobacco germplasm, molecular markers, genetic maps and some useful tools. Using the publically available sequences, TGB developed abundant molecular markers of PIP (Potential Intron Polymorphism) and SSR (Simple Sequence Repeat). For displaying applications with molecular markers, TGB collected Genetic Map and Genetically Modified. In order to make convenience to all researchers, TGB also provided the service BLAST and Primers Developed online.
The Tobacco Genetic and Breeding database at Shandong Agricultural University collects, reproduces and preserves seeds of tobacco and its related species. TGB database will be updated every six months.
Molecular Marker DB
PIP Marker
SSR Marker
EST-SSR
Genomic-SSR
Material DB
Germplasm
Sequence
EST
Genome
TGB
Tobacco Genetics and Breeding Database
Application DB
Genetic Map
Genetically Modified
Tools
BLAST
Primers Development
SSR
PIP
The data in this database are freely available for all academic and non-commercial users. Commercial users are kindly requested to contact us for details of possible licensing arrangements.
Please mail your comments and suggestions to:Long Yang.
Copyright©2011 : Tobacco Laboratory,Shandong Agricultural University PostCode: 271018
C
SSR Develop
This tool finds all simple sequence repeats detail information and designs SSR markers on the appropriate sequences of any species.
A sample
Please enter your query sequences in FASTA format.
Submit
Clear
SSR Develop
D
This tool finds splicing site and potential intron information, and then develops PIP markers based on the appropriate expressed sequences submitted by users.
A sample
Please enter your query sequence in FASTA format.
Submit
Clear
PIP Ddevelop

complete genome has not been finished. However, the exon–intron structure is highly conserved among homologous genes from different plants. This means that the intron positions can be predicted in EST sequences by comparing the tobacco sequences with the genomic sequences of the model plant, *Arabidopsis*. Primer pairs are then designed on both sides of each intron position, and these primers are named PIP markers for detecting ILP (intron length polymorphism) and SNP (single nucleotide polymorphism) in introns (Yang et al. 2007). A pair of primers were designed using the program Primer3 (Rozen and Skaletsky 2000) with 60 bp on each side of the alternative splice joint positions. After testing the designed primers using electronic PCR (e-PCR) and filtering, a total of 12,388 PIP markers was obtained. Compared with the 843 PIP markers for tobacco developed in the PIP database (Yang et al. 2007), this represents a substantial achievement for this type of molecular marker in tobacco. These markers were named with the two-letter abbreviation of the Latin name for tobacco (e.g. NT for *Nicotiana tabacum* and NS for *Nicotiana sylvestris*), followed with 'P' (standing for 'PIP') and a unique number (e.g. NTP0001). Users can key in the ID, PlantGDB ID, or gene name in the PIP query page. Then, by clicking the 'Submit' button, a page of elementary results can be viewed. More information for each marker can be obtained by clicking its name, or linking to the PlantGDB web site using the gene name.

Moreover, users can find out whether the introns exist or not in the querying sequences by using the PIP Develop page. If there are introns in the sequences, then the primers can be automatically developed by the program on this page (Fig. 2).

SSR markers

SSR markers have been developed for maize and rice based on the construction of genomic libraries (Adetimirin et al. 2008; Van Inghelandt et al. 2010; Wei et al. 2009). Compared with traditional methods, the TGB database has adapted a simple channel to identify SSRs from tobacco sequences.

A pipeline in Perl script was used to find the SSR loci and then to develop the SSR markers. In order to find as many SSR units as possible, the parameters were set with simple sequence repeat lengths of 14–20 bp. The procedure consisted of three steps

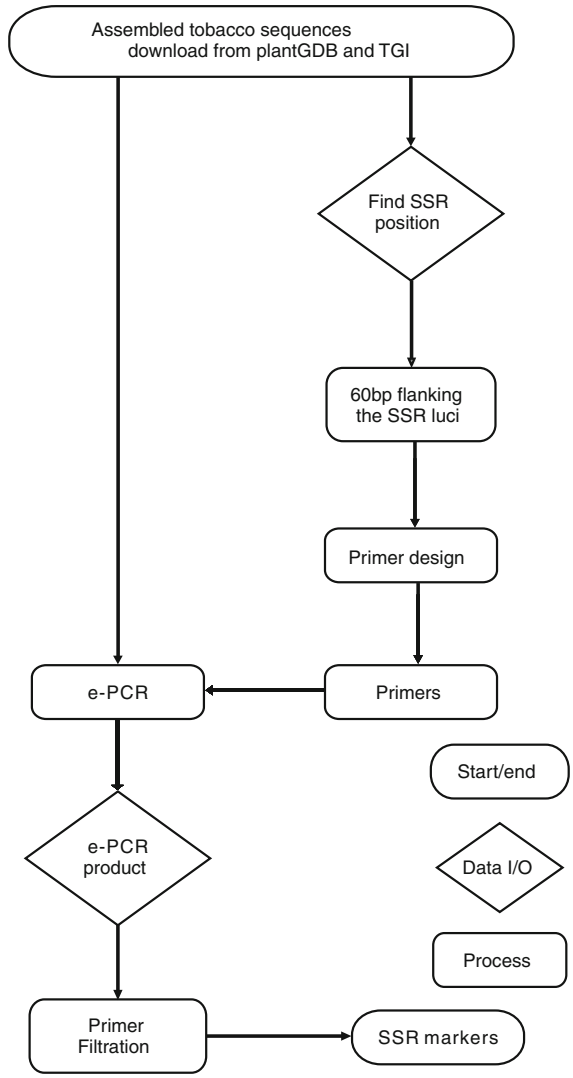


Fig. 3 Flowchart for developing SSR markers. (1) Searching and identifying the SSR loci in all tobacco sequences using a Perl program, with dinucleotide to heptanucleotide motifs. (2) Designing primers using the software ePrimer3 based on sequences with 60 bp on each side of the target loci. (3) Testing the designed primers using e-PCR on all tobacco sequences. Finally, an expected EST–SSR and genomic-SSR marker becomes available after primer filtration

(Fig. 3). The first step was to search and identify the SSR loci in all tobacco sequences using a Perl program, with the motifs from dinucleotides to heptanucleotides. The second step was to design primers, using the software Primer3 (Rozen and Skaletsky 2000) based on the selected EST and genomic sequences with 60 bp on each side of the target loci. The third step was to test the designed

Table 2 Sources of sequences and molecular markers in the TGB database

Name	Total assemblies	PIP	ESR-SSR	Genomic-SSR	Resource
N. tabacum (EST)	131,942	8,233	8,712		PlantGDB
N. sylvestris (EST)	7,612	699	81		PlantGDB
N. benthamiana (EST)	25,297	2,676	1,524		PlantGDB
N. langsdorffi × *N. sanderae* (EST)	6,719	790	234		PlantGDB
TGI (genomic sequences)	1,223,537			66,297	TGI

primers using e-PCR (Schuler 1997) on all tobacco sequences of the four *Nicotiana* species mentioned above. Finally, the best-hit EST–SSR and genomic-SSR markers were made available through the TGB database after filtering the results of e-PCR. Using the above method, a total of 10,551 EST-SSR and 66,297 genomic-SSR markers was obtained, which was much greater than the 5,119 SSR markers developed by Bindler et al. (2011). Moreover, the TGB database provides a platform for developing SSR primers based on querying sequences from any online users (Fig. 2).

There are two types of query pages for SSR markers (EST–SSR and genomic-SSR) in the TGB database. All of the SSR markers were named with the two-letter abbreviation of the Latin name for tobacco (e.g. NT for *Nicotiana tabacum*) followed by the SSR type (e.g. ES for EST–SSR, GS for genomic-SSR) and a unique number (e.g. NSES00001, NTGS00001). Users can obtain detailed information about the two types of SSR markers by clicking the 'Submit' button on the search pages.

The applications database in TGB

Genetic maps

A linkage map is a genetic map of a species or experimental population that shows the position of its known genes or genetic markers relative to each other in terms of recombination frequency, rather than as specific physical distance along each chromosome. The genetic map can be used for tobacco breeding and for further study of the tobacco genome. The TGB database hosts a high-density genetic map created using SSR markers. This genetic map was generated using a F2 mapping population derived from the inter-varietal cross, Hicks Broad × Red Russian. A total of 2,317 SSR and 2,363 loci have been mapped with an average distance of less than 1.5 cM (Bindler et al. 2011). This is a high-resolution genetic map and could be an important tool for use in tobacco breeding research and further analysis of the *Nicotiana* genome.

The GM tobacco study

As a model plant for GM studies, tobacco was the first plant to be genetically modified in 1983 (Hoekema et al. 1983). Some recent studies have showed that GM tobacco could produce a wide range of biologically active proteins and enzymes (James et al. 2000; Magnuson et al. 1998). In order to provide researchers with a similar reference platform for genetic modification, the TGB database collected and unscrambled some interrelated GM research in tobacco. These messages were sorted regularly, and a flowchart of GM tobacco research and a common protocol was created.

Downloading and updating

The TGB database will be continuously updated with new records to keep it current. It provides many germplasm resources that can be used for many different purposes, such as disease resistance and quality improvement. It allows information on the tobacco genome, molecular markers (PIP, EST-SSR and genomic-SSR), GM, genetic maps and germplasms to be downloaded freely by all academic users (Table 2). Using these markers and germplasms, scientists can conduct and design experiments (e.g. gene mapping or quantitative trait loci, gene cloning and quantitative analysis of certain *Nicotiana* species). Researchers are expected to submit to the database program any useful information (e.g. polymorphisms and map position, PCR conduction, characters for a certain tobacco species and GM information) in exchange for information from the TGB database. The submitted information will make the TGB database more valuable and convenient for other researchers.

Conclusions

TGB is a relatively new database developed for *Nicotiana* genetics and breeding studies. It provides many germplasm resources for many different purposes, such as disease resistance and quality improvement. As the TGB program has developed more molecular markers for tobacco than any other program, it can have a positive impact on tobacco genome sequencing and *Nicotiana* studies at the molecular level. In addition, the TGB database not only offers a quick query facility for PIP and SSR markers, but also provides several programs for designing primers online. With the application of genetic maps, GM information and other useful programs, the TGB database will play a significant role in tobacco genetics and breeding studies.

Availability and requirements

The TGB database can be freely accessed at http://yancao.sdau.edu.cn/tgb via the World-Wide Web. A reliable data management system has been developed and all newly released information will be updated on this website. Enquiries concerning the database should be directed by email to lyang@sdau.edu.cn.

Acknowledgments This work was supported by grants from the National Natural Science Foundation of China (No. 30900780), China Postdoctoral Science Foundation (No. 20090461260 & No. 201104647) and Special Funds for Postdoctoral Innovative Projects of Shandong Province (No. 201003030). The authors thank Dr. Jiajie Wu from the College of Agronomy, Shandong Agricultural University, for helpful suggestions on the manuscript.

Conflict of interest The authors declare that they have no competing interests.

References

Adetimirin VO, Vroh-Bi I, The C, Menkir A, Mitchell SE, Kresovich S (2008) Diversity analysis of elite maize inbred lines adapted to West and Central Africa using SSR markers. Maydica 53:143–149

Bai D, Reeleder R, Brandie JE (1995) Identification of two RAPD markers tightly linked with the *Nicotiana debneyi* gene for resistance to black root rot of tobacco. Theor Appl Genet 91:1184–1189

Bindler G, Plieske J, Bakaher N, Gunduz I, Ivanov N, Van der Hoeven R, Ganal M, Donini P (2011) A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development. Theor Appl Genet 123(2):219–230

Day-Rubenstein K, Heisey P (2003) Plant genetic resources: new rules for international exchange. Amber Waves 1(3): 23–29

Denduangboripant J, Setaphan S, Suwanprasart W, Panha S (2010) Determination of local tobacco cultivars using ISSR molecular marker. Chiang Mai J Sci 37(2):293–303

Edwards KD, Bombarely A, Story GW, Allen F, Mueller LA, Coates SA, Jones L (2010) TobEA: an atlas of tobacco gene expression from seed to senescence. BMC Genomics 11:142

Ganapathi TR, Suprasanna P et al (2004) Tobacco (*Nicotiana tabacum* L.)—a model system for tissue culture interventions and genetic engineering. Indian J Biotechnol 3:171–184

Goodspeed TH, Thompson MC (1945) Cytotaxonomy of Nicotiana. Bot Rev 11:533–585

Goodspeed TH, Thompson MC (1959) Cytotaxonomy of Nicotiana II. Bot Rev 25:385–415

Hawkes JG (1977) The importance of wild germplasm in plant breeding. Euphytica 26:615–621

Hoekema A, Hirsch PR, Hooykaas PJJ, Schilperoort RA (1983) A binary plant vector strategy based on separation of vir- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. Nature 303:179–180

James EA, Wang C, Wang Z, Reeves R, Shin JH, Magnuson NS, Lee JM (2000) Production and characterization of biologically active human GM-CSF secreted by genetically modified plant cells. Protein Exp Purif 19:131–138

Knapp S, Chase MW, Clarkson JJ (2004) Nomenclatural changes and a new sectional classification in Nicotiana (Solanaceae). Taxon 53(1):73–82

Lawrence CJ, Dong Q, Polacco ML, Seigfried TE, Brendel V (2004) MaizeGDB, the community database for maize genetics and genomics. Nucleic Acids Res 32(Database issue):D393–D397

Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR (2008) The ups and downs of genome size evolution in polyploid species of Nicotiana (Solanaceae). Ann Bot 101(6):805–814

Lewis RS (2011) *Nicotiana*. In: Kole C (ed) Wild crop relatives: genomic and breeding resources, plantation and ornamental crops. Springer, Berlin

Magnuson NS, Linzmaier PM, Reeves R, An G, HayGlass K, Lee JM (1998) Secretion of biologically active human interleukin-2 and interleukin-4 from genetically modified tobacco cells in suspension culture. Protein Expr Purif 13:45–52

Martz F, Maury S, Pincon G, Legrand M (1998) cDNA cloning, substrate specificity and expression study of tobacco caffeoyl-CoA 3-*O*-methyltransferase, a lignin biosynthetic enzyme. Plant Mol Biol 36(3):427–437

Moon HS, Nicholson JS, Lewis RS (2008) Use of transferable *Nicotiana tabacum* L. microsatellite markers for investigating genetic diversity in the genus Nicotiana. Genome 51:547–559

Moon HS, Nicholson JS, Heineman A, Lion K, Hoeven R, Hayes AJ, Lewis RS (2009a) Changes in genetic diversity of U.S. flue-cured tobacco germplasm over seven decades of cultivar development. Crop Sci 49:498–509

Moon HS, Nifong JM, Nicholson JS, Heineman A, Lion K, Hoeven R, Hayes AJ, Lewis RS (2009b) Microsatellite-based analysis of tobacco (*Nicotiana tabacum* L.) genetic resources. Crop Sci 49:2149–2159

Ravisankar H, Sarala K, Krishnamurthy V, Rao RVS (2008) A software system for tobacco germplasm data. Plant Genet Resour Charact Util 7(2):139–141

Ren N, Timco MP (2001) AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild Nicotiana species. Genome Res 44:559–571

Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, Miller N, Mueller LA, Mundodi S, Reiser L, Tacklind J, Weems DC, Wu Y, Xu I, Yoo D, Yoon J, Zhang P (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Res 31(1):224–228

Ribaut JM, Hoisington D (1998) Marker-assisted selection: new tools and strategies. Trends Plant Sci 3:236–239

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

Ruan CJ, Xu XX, Shao HB, Jaleel CA (2010) Germplasm-regression-combined (GRC) marker-trait association identification in plant breeding: a challenge for plant biotechnological breeding under soil water deficit conditions. Crit Rev Biotechnol 30(3):192–199

Rushton PJ, Bokowiec MT, Laudeman TW, Brannock JF, Chen X, Timko MP (2008) TOBFAC: the database of tobacco transcription factors. BMC Bioinform 9:53

Schuler GD (1997) Sequence mapping by electronic PCR. Genome Res 7(5):541–550

Upadhyaya HD, Gowda CLL, Sastry DVSSR (2008) Plant genetic resources management: collection, characterization, conservation and utilization. Int Crops Res Inst Semi-Arid Trop 6:1–16

Van Inghelandt D, Melchinger A, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120(7):1289–1299

Wei X, Yuan X, Yu H, Wang Y, Xu Q, Tang S (2009) Temporal changes in SSR allelic diversity of major rice cultivars in China. J Genet Genomics 36(6):363–370

Wilde F, Korzun V, Ebmeyer E, Geiger HH, Miedaner T (2007) Comparison of phenotypic and marker-based selection for Fusarium head blight resistance and DON content in spring wheat. Mol Breed 19:357–370

Yang L, Jin G, Zhao X, Zheng Y, Xu Z, Wu W (2007) PIP: a database of potential intron polymorphism markers. Bioinformatics 23(16):2174–2177

Zhang XZ, Kong FS, Yang MK (1990) A design scheme for a database system of crop germplasm resources. Sci Agric Sin 23:79–87