RTPrimerDB: the Real-Time PCR primer and probe database

Filip Pattyn, Frank Speleman, Anne De Paepe and Jo Vandesompele*

Centre for Medical Genetics, Gent University Hospital, 1K5, De Pintelaan 185, B-9000 Gent, Belgium

ABSTRACT

The real-time polymerase chain reaction (PCR) methodology has become increasingly popular for nucleic acids detection and/or quantification. As primer/probe design and experimental evaluation is time-consuming, we developed a public database application for the storage and retrieval of validated real-time PCR primer and probe sequence records. The integrity and accuracy of the data are maintained by linking to and querying other reference databases. RTPrimerDB provides free public access through the Web to perform queries and submit user based information. Primer/probe records can be searched for by official gene symbol, nucleotide sequence, type of application, detection chemistry, LocusLink or Single Nucleotide Polymorphism (SNP) identifier, and submitter's name. Each record is directly linked to LocusLink, dbSNP and/or PubMed to retrieve additional information on the gene/SNP for which the primers/probes are designed. Currently, the database contains primer/probe records for human, mouse, rat, fruit fly and zebrafish, and all current detection chemistries such as intercalating dyes (SYBR Green I), hydrolysis probes (Taqman), adjacent hybridizations probes and molecular beacons. Real-time PCR primer/probe records are available at http://www.realtimeprimerdatabase.ht.st.

INTRODUCTION

Since 1993, when the first polymerase chain reaction (PCR) based DNA amplifications were on-line monitored (1), the real-time PCR technology has successfully evolved to a reliable and easy to use standard method for detection and/or quantification of nucleic acid sequences. PCR product accumulation is measured by a variety of fluorescent detection chemistries, such as generic binding dyes, or sequence specific probes like Taqman (including Minor Groove Binder probes) (2), adjacent hybridization probes (3), or molecular beacons (4). The reliability and accuracy of this method lies in part on the specificity and reaction conditions of the primers and probe(s). While there exists a large set of software tools and Internet sites to design real-time primers and probes, time-consuming and costly optimization experiments are required to ensure that the designed oligonucleotides perform well. Especially when working with highly homologous sequences (e.g. gene families), it might prove difficult to design specific working primers and probes for a particular transcript. We noticed that many researchers experience the same problems when designing real-time PCR primers and probes.

Furthermore, the parallel design of different primers for the same sequence in different laboratories is often a waste of resources, and excludes a certain level of standardization and uniformity. As no specific database for this kind of information is available, we initiated the development of a repository to catalogue validated primer and probe sequences for real-time PCR.

DETAILED INFORMATION

Implementation

RTPrimerDB is implemented on an Oracle 8i relational database management system. The Web application to query and manage the database is based on PHP scripts using the Oracle Call Interface (OCI). The complete application runs on an Apache web server in a Windows 2000 environment.

Data submission

We encourage real-time PCR users to submit their own, validated, primer and probe oligonucleotides. Electronic data submissions are possible after free registration. During registration, personal submitter details are provided, after which an email is sent with the registrar's login name and a temporary password. By changing this password to a more convenient one, the registration is complete and the user can log in to the system and submit custom primer sets. For submission of large datasets, please contact us at administrator@realtimeprimerdatabase.ht.st.

New primer/probe records should contain the official gene name or Single Nucleotide Polymorphism (SNP) identifier (5), the species name and the application for which the primers are used, the detection chemistry, the nucleotide sequences of the primers and the probe(s), and the annealing temperature of the primer pairs. Furthermore, each record provides the

*To whom correspondence should be addressed. Tel: +32 9 2405518; Fax: +32 9 2404970; Email: joke.vandesompele@rug.ac.be

possibility to add submitter's remarks. Currently, only data submissions for primers/probes for human, fruit fly, mouse, rat or zebrafish are allowed, as for these organisms proper controls with respect to the accuracy of the gene name fields are available via LocusLink (6) and the nomenclature databases for these organisms: Genew (Human Gene Nomenclature Database) (7) and GDB (8) for human, MGD (Mouse Genome Database) (9) for mouse, Ratmap (10) and RGD (Rat Genome Database) (11) for rat, FlyBase (12) for fruit fly and ZFIN (Zebrafish Information Network) (13) for zebrafish. Doing so, the presence of aliases or synonyms for official gene names is eliminated in the database. Finally, the possibility to link the PubMed ID of an article where the use of a primer set is discussed makes the primer/probe data more trustworthy.

Querying for primers and probes

The on-line search engine makes it possible to query the database in different ways by type of application, organism, official gene name, detection chemistry, primer or probe sequence, LocusLink ID, SNP identifier, or submitter's name. Search results are listed as a summary of links to the details about specific primer/probe records, or as a complete list of all the details. Each primer and probe set has a unique RTPrimerDB identifier to access them directly or refer to in a publication.

Data integrity

To guarantee accurate data, the sequences in the database are analyzed on regular intervals by BLAST search against the RefSeq database records (6). Upon detection of possible sequence errors, the responsible submitter is contacted by email.

Confidentiality and security

To protect the database information and submitters' passwords, we choose an Oracle database management system to manage the access rights to the different tables and the submitters' passwords.

FUTURE PLANS

In the coming months, we will migrate the data to an Oracle 9i database management system and update the web server with Secure Socket Layer (SSL) support. This will provide improved security when login names and passwords are transmitted. As the system allows easy inclusion of additional applications, sequenced organisms, or newly developed detection chemistries, we will extend the database with the aforementioned items as soon as these become available or if a reference database exists to guarantee the data integrity of the database. In addition, the application will be accessible via a new, permanent URL. The current redirect address http://www.realtimeprimerdatabase.ht.st will continue to operate to provide a link to the new URL.

There is a link to a FAQ section on the RTPrimerDB website where further documentation is available on the use of the database. For additional support or information regarding the database, please write to help@realtimeprimerdatabase.ht.st.

REFERENCES

1. Higuchi,R., Fockler,C., Dollinger,G. and Watson,R. (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N Y)*, **11**, 1026–1030.
2. Livak,K.J., Flood,S.J., Marmaro,J., Giusti,W. and Deetz,K. (1995) Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl.*, **4**, 357–362.
3. Wittwer,C.T., Herrmann,M.G., Moss,A.A. and Rasmussen,R.P. (1997) Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques.*, **22**, 130–131, 134–138.
4. Tyagi,S. and Kramer,F.R. (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.*, **14**, 303–308.
5. Sherry,S.T., Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
6. Pruitt,K.D. and Maglott,D.R. (2001) RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.*, **29**, 137–140.
7. Wain,H.M., Lush,M., Ducluzeau,F. and Povey,S. (2002) Genew: the human gene nomenclature database. *Nucleic Acids Res.*, **30**, 169–171.
8. Letovsky,S.I., Cottingham,R.W., Porter,C.J. and Li,P.W. (1998) GDB: the Human Genome Database. *Nucleic Acids Res.*, **26**, 94–99.
9. Blake,J.A., Richardson,J.E., Bult,C.J., Kadin,J.A. and Eppig,J.T. (2002) The Mouse Genome Database (MGD): the model organism database for the laboratory mouse. *Nucleic Acids Res.*, **30**, 113–115.
10. Steen,R.G., Kwitek-Black,A.E., Glenn,C., Gullings-Handley,J., Van Etten,W., Atkinson,O.S., Appel,D., Twigger,S., Muir,M., Mull,T. *et al.* (1999) A high-density integrated genetic linkage and radiation hybrid map of the laboratory rat. *Genome Res.*, **9**, AP1–8, insert.
11. Twigger,S., Lu,J., Shimoyama,M., Chen,D., Pasko,D., Long,H., Ginster,J., Chen,C.F., Nigam,R., Kwitek,A. *et al.* (2002) Rat Genome Database (RGD): mapping disease onto the genome. *Nucleic Acids Res.*, **30**, 125–128.
12. The Flybase Consortium (2002) The FlyBase database of the Drosophila genome projects and community literature. *Nucleic Acids Res.*, **30**, 106–108.
13. Sprague,J., Doerry,E., Douglas,S. and Westerfield,M. (2001) The Zebrafish Information Network (ZFIN): a resource for genetic, genomic and developmental research. *Nucleic Acids Res.*, **29**, 87–90.