GeneFisher – Software Support for the Detection of Postulated Genes

Robert Giegerich and **Folker Meyer** and **Chris Schleiermacher**[*]
Postfach 100 131, D 33501 Bielefeld, Germany
Phone +49 521 1062911, Fax +49 521 1062962
E-mail: {folker, robert}@techfak.uni-bielefeld.de, chris@genetik.uni-bielefeld.de

Abstract

When a family of genes from closely related organisms is known, there is a certain chance to extract the corresponding gene from the genome of another related organism. This can be done by polymerase chain reaction, provided that a pair of suitable primers can be designed. In contrast to primer design for a single, known target sequence, systematic primer design for an unknown target given a group of homologues can by no means be done manually. *GeneFisher* is a software tool which automates this task, and takes special care to make the impact of the manifold design parameters transparent to the user.

Keywords: PCR, gene extraction, primer design, degenerate primers, WWW

Motivation: Towards Algorithmic Support for Gene Fishing

Standard PCR The polymerase chain reaction (PCR, (Saiki *et al.* 1985; Mullis 1990)) is capable of amplifying a minimal quantity of a specific strand of DNA, such that it can be subjected to standard laboratory techniques like sequencing or restriction profile analysis. Moreover, the exponential amplification power together with the high specifity of the PCR allows it to detect the presence of the target DNA in a complete genome, or in a mixture of DNA from various organisms. This has made PCR one of today's most important techniques in genetics, biotechnology and molecular medicine. The basic pattern of DNA amplification has furthermore exhibited an amazing versatility, giving rise to a large number of derived methods like long PCR, quantitative PCR, competitive PCR, and more (Ferre 1992; Newton & Graham 1994).

The success of a PCR experiment crucially depends on a pair of primers — short oligonucleotides that hybridize to selected templates on complementary strands of the target DNA. One of the difficulties with PCR experiments is its sensitivity to contamination: Assume the goal is to amplify some bacterial gene t in $C.$ *glutamicum*, with primers specifically designed to it. If our solution is contaminated with (say) DNA from *E. coli*, this DNA is likely to contain the gene t' more or less homologuous to t. If the primer templates of t are well conserved in t', then t' will be amplified as well, irrespective of the differences of t and t' between the primer sites. Our PCR run will produce a mixed output, and normally, this is a problematic result.

Precisely this effect of capturing unknown targets is the basis of the method described below.

Gene Fishing via PCR "Gene fishing" refers to the technique where PCR is used to extract a postulated but unknown target sequence from a pool of DNA. The problem is stated as follows:

A family of closely related organisms $o_1, ..., o_n$ exhibits a certain gene[1] with a certain degree of variation. The sequences $t_1, ..., t_n$ are available. The hypothesis is that another related organism o_{n+1} exhibits another variant of this gene, with the unknown target sequence t_{n+1}. How can we (dis)prove this hypothesis, and, in case of proof, determine the target sequence t_{n+1}?

The basic idea is to use $t_1, ..., t_n$ to design a pair of primers that is likely to hybridize to t_{n+1}. If so, running PCR with these primers on the genome of o_{n+1} will extract and amplify t_{n+1}, which can then be subjected to the procedures of sequencing and function validation.

This method has been applied successfully several times. The isolation of a gene from the feline herpes virus by Nunberg *et. al* (Nunberg *et al.* 1989) serves as an example.

But so far, suitable primers had to be guessed, or were determined by trial and error, experimenting with a large number of different primers. In order to turn gene fishing into a systematic technique with a high rate of success, we need algorithmic support for this generalized primer design problem. The software tool *GeneFisher* introduced in this paper is a first step towards a solution.

Unsuccessful Raids As with true fishing raids, there are two different reasons of failure for the PCR experiment described above:

1. The fish is not in the pond: the organism o_{n+1} does not exhibit the postulated gene, and quite rightfully, we obtain no PCR product.

2. The fish is in the pond, but not attracted by our bait: there is a gene t_{n+1}, but the primer template positions determined from $t_1, ..., t_n$ are not well conserved in it. Although the postulated gene exists, we obtain no PCR product.

Alike the fisherman's wife whose husband returns without capture, our method cannot distinguish true from false negatives (case 1. vs. case 2. above). This is inherent in gene fishing, as we are stalking an unknown target. The responsibility of the primer design software is to prevent false negatives whenever the data permit this.

Potential of Systematic Gene Fishing The impact of improving the success rate of gene fishing lies at hand: So far, phylogenetic relationships have been constructed for gene families which were determined in various unrelated sequencing efforts. If gene fishing can be turned into a reliable method, we can systematically construct and complete gene families, targetting on those organisms from which we expect the most significant clues with respect to phylogenetic relation and paths of evolution. This corresponds to a research program postulated by C. Jones in (Jones 1995).

Structure of this Paper In Section 2 we summarize the state of the art in primer design for a single, known target sequence, both from the biological and the algorithmic perspective. Section 3 explains the new problems that arise when designing primers for unknown targets from a family of related sequences. Section 4 introduces the *GeneFisher* tool, our first software solution to this problem. In Section 5, we sketch a method to evaluate the suitability of a particular sequence family with respect to our method. Section 6 concludes by discussing various aspects in which *GeneFisher* can be refined in the future. Finally, we ask for feedback from the research community.

---

[1] Although the name of the method is derived from its application to genes, note that it can be applied to arbitrary strands of DNA.

State of the Art in Primer Design for Known Targets

Biological Criteria for Primer Selection

The aim of almost every PCR experiment is to produce one single fragment of DNA. Primers or even more importantly primer pairs that reliably produce such a single fragment are largely responsible for the success of a PCR experiment.

In order to qualify as primers for PCR experiments, DNA-Oligonucleotides must posses a number of properties. If one or more of the properties are not present in any given nucleotide, the chances for a successful PCR amplification decrease dramatically. The following table gives the required qualities of primers and primer pairs:

- single primer annealing site,
- no secondary structure in each individual primer,
- no primer-primer dimere structures,
- correct primer annealing temperature,
- primer GC content matches target organism,
- primer length,
- length of amplified region,
- sufficent quality of the 3' clamp.

Some of the above criteria are independent of one another, others like melting temperature and GC content, or melting temperature and number of possible annealing sites are dependent on each other.

The more of the above criteria a primer (pair) fulfills, the higher are the chances for a successful amplification of the desired region. No generally accepted consensus exists among molecular biologists as to what parameters will yield the best results in a PCR experiment. The large number of areas the PCR is used in and the diversity of target sequences that are used as matrices lead to the conclusion that there is no single optimal setting, but rather a range of parameters for different settings.

Property	Value Source
GC content	45-55 % (Lowe *et al.* 1990) like target organism (Pershing 1993)
3' terminal base	2 GC residues (Lowe *et al.* 1990) 1 Thymin (Kwok *et al.* 1990)
3' clamp GC content	as high as possible (Lowe *et al.* 1990) not too high (Pershing 1993; Rychlik 1995)
ampl. region length	max 600 bp (Lowe *et al.* 1990) up to 50 kbp (Pershing 1993)
possible degeneracy	up to 1024 (Innis *et al.* 1990) 516 (Compton 1990) 200 (Lowe *et al.* 1990)
stability of the 3' clamp	as low as possible (Rychlik 1995) (see 3' GC content)

Table 1: Recommended Criteria for Primer Selection

Table 1 shows some of the parameters suggested in PCR literature; it also shows the lack of consensus among researchers as to what constitutes a good primer.

Existing Software A number of software products exists that facilitate primer selection for PCR experiments. These programs offer a wide variety of services and use different data to compute the primers.

Table 2 lists the PCR primer design products that we are aware of, with information about the criteria they implement and the type of input data used:

Some tools implement only a fraction of features and rely on external programs to compute some properties (e.g. computation of melting temperature, T_m), others can only cope with a limited amount of data (most PC/DOS based programs). The most complete tool in this list is probably Oligo (see Table 2).

The fact that still many biologists use a manual approach when designing degenerate primers stems (at least to some degree) from the shortcomings in functionality, comprehensability and user guidance of the tools above. Some biologists claim that transparency is

Name	Platform	GUI	Input	Primer-design	T_m
Oligo	MS-Dos	+	1-DNA	+	+
Pgen	MS-Dos		Protein	+	+
Primer (Stanford)	Sparc SunOS 4.X	-	1-DNA	-	-
Primer (Whitehead)	C-source	-	1-DNA	+	-
Amplify	Macintosh	+	?	?	-
OSP	Sun,Vax	-	1-DNA	+	+
Primers	MS-Dos	-	1-DNA	+	+
Nuc-It	MS-Dos	?	1-DNA	+	+
Primer Detective	MS-Dos	?	1-DNA	+	+
DesignerPCR	Windows	+	1-DNA	+	+
Degen-Designer	NeXT	+	1-DNA	+	+
PrimeGen	?	-	1-DNA	-	?
Primer-Master	MS-Dos	-	1-DNA	+	?

Table 2: Primer Design Software

their prime motive for using a 'hands on' approach instead of any software tool. Transparency is a very important feature when it comes to flexible, user-changeable parameters: the user has to be informed of the effects which the chosen settings have on primer selection.

Additionally, seasoned PCR experts are also critical of using any type of software tool because it limits their range of choices available. So the challenge is to increase functionality and flexibility at the same time with transparency of primer design software.

Primer Design from Sequence Homologues for Unknown Targets

Based on the assumption that genes with related function (or the same gene in two different organisms) show a high degree of sequence homology, the idea arises to design primers from a multiple sequence alignment. This technique is already in widespread use among molecular biologists, but so far there has been no software supporting this more general case of primer selection.

When using a sequence family instead of a single sequence to compute the primers, a

number of new problems arise:

- primers have to be extracted from a multiple sequence alignment;
- degenerate primers have to be built, to cope with variant positions in the consensus;
- properties of degenerate primers have to be calculated;
- primer degeneracy has to be limited.

Degenerate primers are multiple copies of an oligonucleotide, where in selected positions, the base is varied. A mixture of all variants is used in the PCR experiment. Dealing with degenerate primers complicates practically all aspects of primer evaluation. Both extracting the primers from a sequence alignment and ascertaining their properties are tasks that are beyond the scope of what can be done without algorithmic support. The effort of checking for possible primer-primer interaction is quadratic in the degree of degeneration. Determining the secondary structure(s) of a 512-fold degenerated oligonucleotide is a daunting task if done without computer help.

This complex process of primer evaluation is governed by a large number of parameters. In order to be practically useful, the primer design software must make good efforts to explain to its user the effect that various selection criteria and their parameters actually have on primer selection. A carefully designed user interface and good feedback mechanisms are absolutely essential.

The *GeneFisher* Software

Functionality Starting from either a set of unaligned sequences or a multiple sequence alignment (Altschul & Lipman 1989), *GeneFisher* first calculates a consensus sequence, then determines possible priming regions for forward and reverse primers. This is done by evaluating each position using a set of criteria individually adjustable by a user within reasonable bounds. Individual primers are subjected to the following criteria. In parenthesis,

we include the defaults used by *GeneFisher*. Please note, that these are just defaults, and not recommendations. Different PCR applications require significantly different parameter settings.

- Global primer properties
  - Melting temperature T_m (55-65 deg. C),
  - Guanine/Cytosine (GC) content (45-65%),
  - Primer size (15-18 bp),
  - Primer degeneracy (8 fold),
  - Uniqueness
    (multiple occurrences of primer homologuous regions are monitored) (1 occurence),
- 3' clamp properties
  - 3' clamp GC content (45-55%),
  - 3' clamp degeneracy (4 fold),
  - 3' terminal residue (IUPAC base 'w', i. e. Adenine or Thymine),
- Structural properties[2]
  - Avoidance of hairpin structures,
  - Patterns (search primers and amplified region for any patterns e.g. restriction sites).

From the best forward and reverse primers, pairs are formed and evaluated with respect to:

- Melting temperature compatibility
- Size of amplified region
- Avoidance of primer-primer interaction

Parameters and Feedback The process of primer selection is governed by 12 independent parameters. Although we try to give reasonable defaults, most likely every user will have to experiment with these values. Figure 1 gives an example.

Following each calculation, *GeneFisher* 'justifies' the number of primers rejected according to the actual parameter settings. In the example of Figure 2, most primers were rejected for not matching the overall GC content setting. In this way the "reject statistics" hint

---

[2]Structural properties do not lead to rejections, but primers are labelled accordingly

to those parameters which should primarily be modified to optimize calculation results. This has been highly appreciated by the laboratory practitioners.

A special case covered by the software is the calculation of primers for a single target sequence. With the implemented criteria *GeneFisher* exceeds the functionality of most other software tools (Pershing 1993; Rychlik 1995; Lowe *et al.* 1990) that are available for this restricted case.
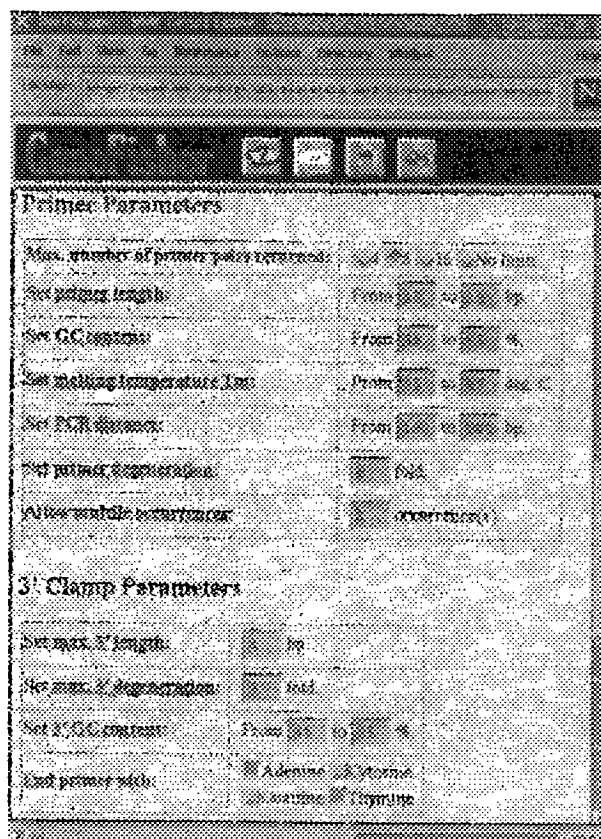

Figure 1: Parameter Input

The WWW Interface Since successful primer design constitutes a quite complex process, adequate user guidance is inevitable. *GeneFisher* accomplishes this by a state of the art World Wide Web (WWW) user interface along with reasonable default settings and a facility to individually modify almost every pa-

rameter specifying primer properties.

Genefisher can be used via WWW at http://www.techfak.uni-bielefeld.de/techfak/ags/pi/GeneFisher/.

The interface extensively uses HTML 3.0 constructs and Netscape's frame feature (Net 1996) and requires Netscape Navigator 2.0 (see Figures 1 and 2). Remote users must enter pre-aligned data.[3] All parameter defaults are shown in the form and can be modified at will. The 'submit' button starts the primer design engine, which returns a link to a newly generated page presenting the results and other feedback information.

All future enhancements of *GeneFisher* will be accessable via the WWW interface. However, the current version of the *GeneFisher* software is also available from the authors for local installation. It offers a more convenient graphical interface. In particular, it provides an extended graphical overview of primer locations, hair pins and other properties (an example is Figure 3).

Internal Modularity We chose a modular approach in designing the software with an eye towards future modifications and extensions. First of all, there is a primer design engine separate from the user interface. It can be combined with the WWW interface, a more comfortable graphics interface or with a primitive command line interface.

To generate primers from raw sequences, *GeneFisher* imports certain external software tools, e.g. a choice of user selectable alignment tools. Currently we are using ClustalW (Higgins & Sharp 1988), but we see good reasons to include other algorithms like MSA (Lipman, Altschul, & Kececioglu 1989) or the divide & conquer approach of (Dress, Fuellen, & Perrey 1995). Thus an improvement of existing programs or the developement of new methods is easily included to optimize the functionality of
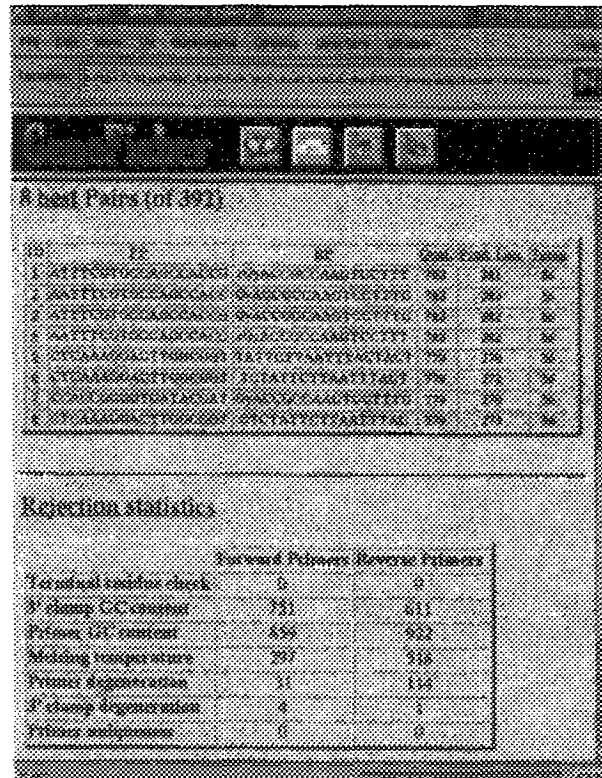

Figure 2: The Reject Statistics

GeneFisher.

The Primer Generation Algorithm

From a computational point of view biological sequences are members of a set of finite sequences, composed from characters over a certain alphabet. To explain the techniques *GeneFisher* implements, the following notation is used:

- $\mathcal{A}$ is the extended genetic or IUPAC code[4] alphabet,
- $\mathcal{A}^*$ is the set of all finite sequences of characters from $\mathcal{A}$,
- s denotes sequences from $\mathcal{A}^*$,
- $_i : s :_j$ denotes the subsequence of s between position i and j.

Starting with a multiple sequence alignment, *GeneFisher* first generates a consensus se-

---

[3] Currently, we lack the server capacity for computing multiple alignments for remote users.

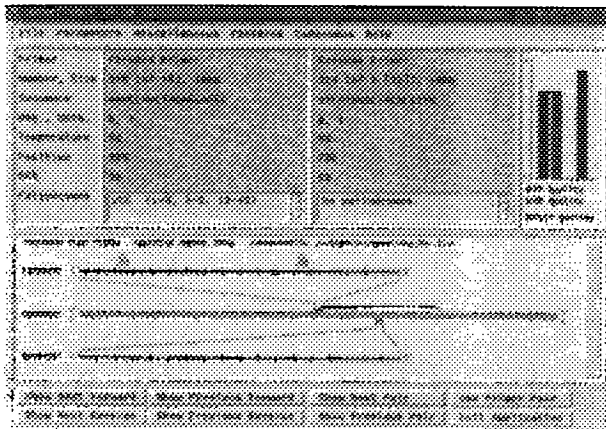[4] abbrev. International Union of Pure and Applied Chemistry (Meyers 1988)

Figure 3: The *GeneFisher* Local Version

quence *s*. Then a test function, actually an implementation of a finite automaton, takes *s* as input, finds suitable substrings $_i$:*s*:$_j$ according to the test criteria mentioned above and outputs primer data. The following pseudo code represents the procedure:

```
FOR Position = SequenceLength DOWNTO MinPrimerLen {
    FOR Len = MaxPrimerLen DOWNTO MinPrimerLen {
        IF P_3'_TERMINAL(Position-Len : s :Position) AND
            P_GENERAL(Position-Len : s :Position)
        THEN AcceptPrimer
        ELSE JumpToNextPrimerPosition
    }
}
```

Primer evaluation proceeds from the 3' to the 5' end, since the former normally underlies stronger restrictions. Assuming that longer primers are more specific in PCR reactions than shorter ones, those are tested first. P_3'_TERMINAL summarizes those criteria which solely refer to the 3' end of the primer (terminal residue, 3' GC content, 3' degeneracy). If either of these fails, the outer loop is advanced. P_GENERAL comprises the other criteria; if these fail, the inner loop is advanced.

The rejection statistics lists the criteria in the order in which they are tested. In the actual version we use a static order of test criteria described below:

1. 3' terminal residue,
2. 3' clamp GC content,
3. primer overall GC content,
4. melting Temperature T_m
5. primer degeneracy
6. 3' clamp degeneracy
7. uniqueness (of binding site)

After this forward pass the consensus sequence is reversed and complemented to generate reverse primers. Finally, each primer from the forward primer pool is matched against each primer from the reverse primer pool to produce pairs using the following criteria:

- size of amplified region (PCR product length),
- avoidance of dimere structures,
- melting temperature compatibility.

Computational Resources Running *GeneFisher* requires a modest amount of computational resources[5]. The asymptotic complexity of the primer design phase is $O(m * n^2 + l * d^2)$, where n denotes the input sequence length, l the maximal primer length, m the allowed variation in primer length and d the degree of degeneracy. As a practical example, *GeneFisher* took about 15 seconds on a SUN Sparc 10 with 32 MB to design primers for the *dap*F gene in *C. glutamicum* using sequences from *Y. pestis*, *P. aeruginosa*, *M. tuberculosis* and *E. coli*.

Evaluation

Evaluation of the Primers generated by *GeneFisher*

The Primers generated by GeneFisher have been tested in various ways.

- They were compared to manually-designed primers for *spo*T from *C. glutamicum* that had already been used in a successful PCR run. *GeneFisher* found all the known primers; 3 were exact and 2 were one base off.

---

[5]This does not include the multiple alignment.

- Primers for the postulated *dap*F gene in *C. glutamicum* were designed by *Gene-Fisher* from the related genes from *E. coli*, *S. vibrio*, *M. leprae* and *P. aeroginosa*. The parameters were successively strengthened untill only 2 primer pairs remained. These primers amplified a fragment from *C. glutamicum* of the desired length and with the expected restriction profile. The precise sequence is yet to be determined.

- We also evaluated *GeneFisher* against the PRIMER program (Lincoln, Daly, & Lander 1995) Primers for a single target sequence s_1 were designed by PRIMER, primers for a family of related genes were designed by *GeneFisher*. For details see Table 3 and explanations below.

Evaluating the Suitability of Input Data

Let us return to the issue of false negatives, already discussed in the introduction. Can we examine the input and inform the user if the sequence input is too inhomogeneous to produce primers for successful genefishing?

One plausible test is based on the idea that it should be possible to catch each of the n given sequences based on the rest.

In the following setup, we use an alternative Primer design program[6] to generate primers for an individual sequences and *GeneFisher* for the multi-sequence case. Let $f = s_1, \cdots, s_n$ be a family of sequences, and $f_i = f \setminus s_i$. Let $\delta(x, y)$ the Levenshtein distance (i.e. the edit distance under the unit cost model, adapted to the degenerate code) of sequence x and y. We denote by P1(s) the primers suggested for the single sequence s and by GF(f) the primers constructed for the sequence family f.

The first test checks the chance to amplify some s_i via primers from f_i. We calculate for each $x \in GF(f_i)$, $quality(x) :=$

---

[6]This is because we have an eye on validating *Gene-Fisher* against previous, simpler programs. Later *GeneFisher* will be employed for the single sequence case, too.

$x \in GF(f_i)$	$y \in P1(s_i)$	quality (x)
TCyAChTGyrCyCyCA	TACTTGTGGCTCCCA	1
TTyTCyAChTGyrCy	CCTTCTCTACTTGTGGCT	1
AArGCyTTyTCyAChTGyr	GGCCTTCTCTACTTGTG	0
AArGCyTTyTCyAChTGy	CAAGGCCTTCTCTACTT	2
rAArGCyTTyTCyAChT	GCAAGGCCTTCTCTA	2
rAArGCyTTyTCyACh	GCAAGGCCTTCTCTA	2
hrAArGCyTTyTCyA	TGCAAGGCCTTCTCT	2
ACwTyywCTGTGABmy	CCCACTTTCTCTGTGATAT	0
yACwTyywCTGTGAbm	CCCACTTTCTCTGTGATA	0
mhyACwTyywCTGTGA	CCCACTTTCTCTGTGA	0
TwyTGGyAThTGCy	ATGTGGCCATCTGCT	3
ATGGChTATACyrmTwy	GGCCTATGACCGCTA	1
ATGGChTATACyrmTw	GGCCTATGACCGCTA	1
GyATGGChTATACyrmT	CCATGGCCTATGACC	3
GyATGGChTATACyrm	CCATGGCCTATGACC	3
GyATGGChTATACyr	CCATGGCCTATGACC	3
bGyATGGChTATACy	CCATGGCCTATGACC	2
ybGyATGGChTATA	GCCATGGCCTATGAC	3

Table 3: Quality of *GeneFisher* primers

i	score(i)
1.	1.61
2.	1,69
3.	1,6
4.	2,2
5.	2

Table 4: Aggregated quality for OR17-40 family

$min\{\delta(x, y), x, y \in P1(s_i)\}$. Table 3 shows some data obtained using the OR17-40 genes.

To evaluate the family f as a whole, we calculate for each choice (s_i, f_i), $score(i) := average \{quality(x) \mid x \in GF(f_i)\}$. Note that for both quality and score, the value 0 is the theoretical optimum.

Table 4 shows these scores for the homogeneous family given above. In Table 5, one unrelated gene was added, which leads to a distribution of scores where the unrelated sequence scores worst. This is a hint to the user with

i	score(i)
1.	1.6
2.	1.2
3.	1.66
4.	2.8
5.	2.8
unrelated	3.25

Table 5: Aggregated quality for OR17-40 family plus one unrelated gene

respect to the suitability of this sequence.

Conclusion

GeneFisher has been used within our local genetics group and in a cooperation with KFA Jülich. The WWW interface is its most recent part, and *GeneFisher* is becoming available for remote use just at the time of writing. Users are kindly asked to provide feedback, both with respect to tool usage, and to the success / failure of PCR runs with primers designed by *GeneFisher*.

We believe that GeneFisher surpasses previous software in functionality and ease of use. But in our long-term plans, GeneFisher as of today is only a starting point. Numerous improvements are necessary to support the complex overall process of gene fishing. Consider the well-known example of the feline herpesvirus as reported by Nunberg et. al. (Nunberg *et al.* 1989). That study started from an alignment on the amino acid level, then used back-translation and codon usage tables to reduce the degeneracy. This is a very powerful technique when homology on the DNA level is lower. We would like to import such functionality into GeneFisher, as we do with multiple alignments.

Our current plans include many algorithmic improvements, some simple, and some rather substantial ones.

- Internally the program is prepared to give the user control over the way the consensus sequence is formed, and the order in which the primer test criteria are applied. However, we are not sure whether it would be wise to offer such flexibility.
- More experience must be gained in the use of the reject statistics to tune the parameters. Certainly, the statistics should also include percentages of rejected primers. In some cases an account of rejected primer pairs might also be helpful.

- Structure calculations should be based on energy models rather than the usual rules of thumb.
- Global alignments have recently been criticized (Sander 1995) because they are geared towards overall consensus. We plan to experiment with algorithms that determine consensus sequences in a more localized way.
- Algorithmically it is tempting to replace the generate and test paradigm by some more intelligent search of the primer design space. From the viewpoint of program flexibility, it may not be wise to do so.

Finally, if GeneFisher proves to be successful as we hope it will, our plan is to work out a systematic strategy for large-scale gene fishing as outlined in the introduction.

Acknowledgements

References

Altschul, S. F., and Lipman, D. J. 1989. Trees, stars, and multiple biological sequence alignment. *J. Appl. Math.* 49:197–209.

Compton, T. 1990. Degenerate primer for dna amplification. *PCR Protocols* 39–53.

Dress, A.; Fuellen, G.; and Perrey, S. 1995. A devide and conquer approach to multiple alignment. *Proceedings ISBM-95* 107–113.

Ferre, F. 1992. Quantitative or semiquantitative pcr: Reality versus myth. *PCR Methods and Applications* 2(1-9).

Higgins, D. G., and Sharp, P. M. 1988. Clustal: A package for performing multiple sequence alignments on a microcomputer. *Gene* 237–244.

Innis, M.; Gelfand, D.; Sninsky, J.; and White, T. 1990. Pcr protocols: A guide to methods and applications.

Jones, S. 1995. An update lesson from whole-genome sequencing projects. *Curr. Opinions. Genet. Devel.* 5:349–353.

Kwok, S.; Kellog, D. E.; McKinney, N.; Spasic, D.; Coda, L.; C., L.; and Sninsky, J. J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acid Research* 18:999–1005.

Lincoln, S.; Daly, M.; and Lander, E. 1995. *Primer v2.2.* Whitehead Institute for Biomedical Research.

Lipman; Altschul; and Kececioglu. 1989. A tool for multiple sequence alignment. *Proc. Natl. Acad. Sci.* 86:4412–4415.

Lowe; Todd; Sharefkin; John; Yang; Yang, S. Q.; Dieffenbach; and W., C. 1990. A computer program for selection of oligonucleotide primers for polymerase chain reactions. *Nucleic Acids Research* 18(7):1757.

Meyers. 1988. *Chemie Duden.* Dudenverlag, Mannheim, Wien, Zurich.

Mullis, K. B. 1990. The unusual origin of the polymerase chain reaction. *Scientific American* 36–43.

Netscape Communications Corporation, http://home.netscape.com/comprod/ products/navigator/version_2.0/index.html. 1996. *Introducing Netscape Navigator 2.0 and Netscape Navigator Gold 2.0.*

Newton, C. R., and Graham, A. 1994. Pcr.

Nunberg, J. H.; Wright; E., C. G.; Petrovski, E. A.; Post, L. E.; T., C.; and Gilbert, J. H. 1989. Identification of the thymidine kinase gene of feline herpesvirus: use of degenerate oligonukletiodes in the polymerase chain reaction to isolate herpesvirus gene homologs. *J. Virology* 63:3240–3249.

Pershing, D. 1993. Target selection and optimization of amplification reactions. *Diagnostic Molecular Biology, Principles and Applications* 88–105.

Rychlik, W. 1995. Selection of primers for polymerase chain reaction. *Molecular Biotechnology* 3:129–134.

Saiki, R. K.; Faloona, F.; Mullis, K. B.; Horn, G. T.; Ehrlich, H. A.; and Arnheim, N. 1985. Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350–1354.

Sander, C. 1995. Speaker topic. Dagstuhl Workshop on Bioinformatics.