

A New Multiplex-PCR for Urinary Tract Pathogen Detection Using Primer Design Based on an Evolutionary Computation Method

Liliana Torcoroma Garcia^{1*}, Laura Maritza Cristancho¹, Erika Patricia Vera¹, and Oscar Begambre²

¹Program of Bacteriology and Clinical Laboratory, Universidad de Santander - UDES, 680003 Bucaramanga, Colombia

²School of Civil Engineering, Universidad Industrial de Santander, 680003 Bucaramanga, Colombia

Received: July 7, 2014
Revised: October 3, 2014
Accepted: June 9, 2015

First published online
June 9, 2015

*Corresponding author
Phone: +57-7-651-6500, Ext 1212;
Fax: +57-7-651-6492;
E-mail: l.torcoroma@udes.edu.co

pISSN 1017-7825, eISSN 1738-8872

Copyright© 2015 by
The Korean Society for Microbiology
and Biotechnology

This work describes a new strategy for optimal design of Multiplex-PCR primer sequences. The process is based on the Particle Swarm Optimization-Simplex algorithm (Mult-PSOS). Diverging from previous solutions centered on heuristic tools, the Mult-PSOS is self-configured because it does not require the definition of the algorithm's initial search parameters. The successful performance of this method was validated *in vitro* using Multiplex-PCR assays. For this validation, seven gene sequences of the most prevalent bacteria implicated in urinary tract infections were taken as DNA targets. The *in vitro* tests confirmed the good performance of the Mult-PSOS, with respect to infectious disease diagnosis, in the rapid and efficient selection of the optimal oligonucleotide sequences for Multiplex-PCRs. The predicted sequences allowed the adequate amplification of all amplicons in a single step (with the correct amount of DNA template and primers), reducing significantly the need for trial and error experiments. In addition, owing to its independence from the initial selection of the heuristic constants, the Mult-PSOS can be employed by non-expert users in computational techniques or in primer design problems.

Keywords: Primer design, multiplex-PCR, urinary tract infections, uropathogens, particle swarm optimization, metaheuristic optimization

Introduction

The gold standard methods for the diagnosis of many infectious diseases caused by bacteria are based on culturing clinical samples in order to carry out a biochemical identification of their etiological agent. These diagnostic tools are widely utilized owing to their affordability, ease of application, and performance. However, these traditional techniques are time-consuming (48 to 72 h) for the complete identification of many species and exhibit uncertainty associated with their lack of sensitivity or specificity (microorganism-dependent). These drawbacks have hindered the etiological identification of many infections in clinical practice, and have led to a tendency towards empirical diagnosis, control, and treatment, especially in the most prevalent infectious diseases.

In the past two decades, the advent of molecular techniques based on DNA amplification (such as polymerase

chain reaction (PCR)), have allowed the gradual replacement of the classic techniques for microbiological identification. Owing to their high sensitivity and specificity, many PCR assays have been widely accepted as the new gold standard for the detection of viruses [2] and fastidious pathogens [11, 13, 16, 22, 30], for antimicrobial resistance profiling [5, 8, 13], and for etiological identification of body fluid infections [2, 26].

The PCR is an *in vitro* replication that allows for the massive and rapid amplification of specific sequences of small DNA fragments [18]. This is a three-step iterative process characterized by the initial hot denaturation of a double-stranded DNA template, followed by an annealing of the specific primers on the target, and finishing with an extension step of the annealed primers by the DNA polymerase reaction. For the massive replication of DNA templates, this process is usually repeated between 25 and 40 times. In order to obtain an efficient *in vitro* replication

in PCR assays, relevant parameters should be coordinated; that is, factors such as cycle conditions (time and temperature of each phase of the cycle and number of cycles) and concentration of the reactants in the mix (primers, magnesium ions, dNTPs, and DNA templates). Nevertheless, the most critical factor in the performance of an *in vitro* replication (uniplex and multiplex) is the adequate selection of the primer sequence [1, 6, 27, 32]. The primer design process in a PCR is complex and time-consuming owing to the multiple criteria that must be taken into account: (i) the primer length, ranging from 18 to 26 bases; (ii) the guanine-cytosine percentage, between 40% and 60%; (iii) the melting temperature (T_m), with values ranging from 45°C to 65°C, with 5°C as the maximum difference between primers; (iv) the absence of base repetitions of more than four bases; and (v) the presence of G, C, GC, or CG in the primer 3' end. Inconsistencies in these criteria can lead to the formation of unspecific products in PCRs [3], as a consequence of inter- and intra-molecular structures that diminish the potential performance, specificity, and/or sensitivity of DNA amplification [6].

In bacterial infections of multiple etiologies, the identification of the causal agent by PCR techniques is expensive and laborious owing to the need for numerous individual analyses. In these cases, assays of PCR able to amplify multiple DNA targets in a single reaction, such as Multiplex-PCR, have been shown to be the most efficient option. In this type of PCR, multiple pairs of primers are included in the same test tube. However, owing to the several parameters to be considered in the optimization process for the Multiplex-PCR conditions, the implementation of this technology for clinical diagnoses represents a great challenge, especially because of the primer design problem. In this technique, each of the primers used in a mix should be designed to satisfy all the aforementioned parameters. Additionally, in order to efficiently obtain the amplification of multiple target sequences, all the primer sets used should be adjusted as if they were a single pair of primers. Furthermore, for the electrophoresis analysis, all the product sizes should be planned to allow the appropriate resolution in the gel, with a minimum difference among the PCR products of 50 bp.

This practice is usually a highly time-consuming and expensive approach as a result of human error. To overcome these difficulties, in the last decades, several heuristic optimization methods have been implemented for the solution of the Multiplex-PCR primer design problem [21, 29]. In this context, genetic algorithms [29, 31] and particle swarm optimization [33] techniques have been recently

used to solve the problem. However, the major drawback of these approaches is the requirement for an *a priori* definition, by the user, of the parameters that control the algorithm performance and the fact that these parameters are problem-dependent.

In order to overcome these challenges, Begambre and Laier [7] proposed a new self-configured evolutionary hybrid algorithm. This algorithm does not require the definition of its initial search parameters, allowing an operator (who is not a linear or non-linear programming expert) to obtain the optimal or quasi-optimal conditions for amplification of multiple DNA fragments. In this work, a new strategy for Multiplex-PCR primer design is presented based on Particle Swarm Optimization (PSO) – Simplex (s), PSOS [7]. The PSOS algorithm was adapted and used as a tool for *in silico* rational design of the primers for the Multiplex-PCR assays. This adaptation is called Mult-PSOS. The efficiency and robustness of this method was evaluated *in vitro* using the Mult-PSOS predicted primers for uropathogen detection. The goal of this multiplex-PCR was to allow, in a single assay, the efficient molecular identification of the most common bacteria associated with urinary tract infection (UTI) (*Escherichia coli*, *Proteus mirabilis*, *Proteus vulgaris*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Staphylococcus saprophyticus*, and *Enterococcus* sp.) [10, 12, 14, 25]. It is worth noting that UTIs are the most common and serious infectious diseases among outpatients and inpatients, and represent a significant healthcare burden [10].

Materials and Methods

The Particle Swarm Optimization and the Simplex Algorithm

As described in Refs. [28], [29], and [31], the manual primer design is unstable and the software available to design primers is not capable of selecting the best candidates. Therefore, in this work, the problem is defined as an optimization problem whose principal objective is to build candidate primers and to choose the best candidate according to a fitness function and its restrictions. For the sake of clarity, in this section, a brief description of these algorithms is presented.

The Simplex Algorithm

Here, we present a short review of the classical version of the algorithm. The method uses a polytope (a polytope (in this case referred to as the Simplex) is a geometric object with plane sides, such as a polyhedron, in three dimensions) with $N+1$ vertices, where N is the number of variables in the optimization problem [19]. The Simplex explores the search space (all the possible values that the variables can take), permitting each of its vertices only

three movements (or actions): expansion, contraction (shrinkage), and reflection [19]. These actions allow the Simplex to explore the topography of the fitness function and finally to determine the optimum solution (minimum point). The Simplex coefficients used in this work were $\sigma = 0.5$; $\varphi = 0.5$; $\chi = 2$; $\rho = 1$ (standard case) as recommended in Ref. [7] with a maximum number of iterations of 200 (stopping criterion).

The PSO Algorithm

PSO is based on a simplified social interaction model. It is known that sharing information in social environments may offer evolutionary advantages to individuals. PSO is a member of the group of computational intelligence methods (or heuristic methods) to solve optimization problems. In this context, the particle (or individual of the swarm) position is updated using social information shared by the swarm members (see Eqs. (a) and (b)) and each particle tries to change its position to a point where the swarm (or the particle itself) has a higher fitness function value [15].

The search is controlled by the following expressions [15]:

$$p_{i+1}^k = p_i^k + vel_{i+1}^k \quad (a)$$

$$vel_{i+1}^k = \omega vel_i^k + C_1 Ran_1(B_i^k - p_i^k) + C_2 Ran_2(B_i^g - p_i^k) \quad (b)$$

where p_i^k is the position of particle i at iteration k (a position vector whose size is equal to the number of variables in the problem), vel_{i+1}^k is the updated velocity of particle k at iteration $i+1$, B_i^k stands for the best position of particle k at iteration i , B_i^g is the best global position at iteration i (the best particle in the swarm), Ran_1 and Ran_2 are independent random numbers, and ω is the inertia weight (a factor that controls the impact of the previous velocity on the current particle velocity). From Eq. (b), it is evident that the factors C_1 (cognitive parameter), C_2 (social parameter), and ω (inertia weight) must be carefully chosen by the user to obtain a good result. The values of the PSO factors employed in this study were

$$0 < \omega < 1.5; C_1 + C_2 \leq 4, \text{ with a population of 10 particles.}$$

Particle Swarm Optimization – Simplex PSOS

To solve the Multiplex-PCR primer design problem, several procedures have been employed. Taking into account that the problem has a huge number of possible solutions (the search space is enormous) and discrete variables, many authors have used heuristic algorithms (*i.e.*, genetic algorithms, particle swarm optimization). The principal advantage of these techniques is that they do not require the computation of derivatives or Hessians (difficult or impossible to obtain in this class of problems). However, heuristic methods need the designation of many users' defined initial factors to control the search (*e.g.*, the cognitive parameter, social parameter, and inertia weight in Eq. (b)). To overcome this drawback, we adopted in this work the concept of a

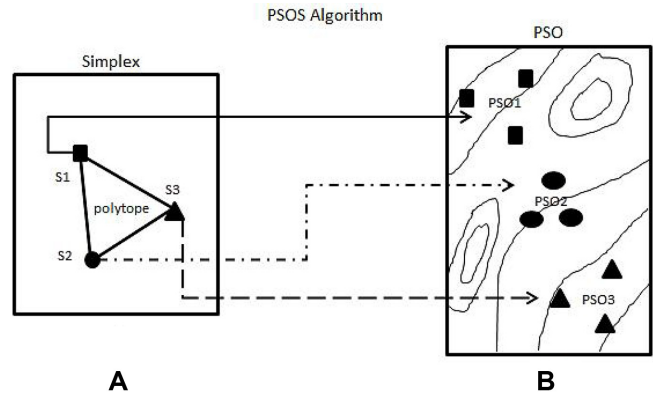


Fig. 1. Basic PSOS heuristic.

(A) Heuristic factor search space and the polytope. S1, S2, and S3 are the Simplex vertex. (B) Fitness function values (given by Eq. (A) or Eq. (B)). PSO1, PSO2, and PSO3 are the swarms controlled by each Simplex vertex.

self-configured algorithm [7]. This form of algorithm combines two heuristic methods. The first algorithm (in this case, the nonlinear Simplex, or Simplex) determines the factors that control the performance of the second algorithm (the PSO) and the PSO performs the search in the fitness function values space. This topology guarantees that, no matter what the initial parameters of the first algorithm are, the second procedure always finds a very good solution without user intervention [7]. Fig. 1 shows the plain PSOS.

Based on the ideas outlined above, the Multiplex-PCR primer design proposed in this work can be outlined in four steps:

- 1) Set the definitions and variables of the Multiplex-PCR optimization problem;
- 2) Select the first primer pair: first optimization;
- 3) Multiplex-PCR optimization process: second optimization;
- 4) Stopping criterion.

The selection of the first primer pair is made using an optimization process. This primer pair, as well as its restrictions, is used for the second optimization process. Both of the two optimization processes are carried out using an adapted PSOS algorithm [7], called Mult-PSOS. For the sake of precision, each step is explained in detail in the next section.

Multi-PSOS Optimization for Multiplex-PCR Primer Design

As described above, the algorithm has four main steps and its final goal is to obtain a Multiplex-PCR optimum primer design.

- 1) **Set the definitions and variables of the Multiplex-PCR optimization problem.** In this study, the definitions described previously in Ref. [28] were applied for the basic parameters in the primer design problem:

Definition of the target sequence or GD: In order to improve the specificity and sensitivity of the Multiplex-PCR for infectious

disease diagnosis, the target sequence or GD should fulfill all of the following criteria: (i) a degree of variation in the sequence sufficiently significant to identify the target organism; (ii) high GC content (between 40% and 60%); (iii) multiple copies of the target genome (if possible); and (iv) low presence of tandem repetitions in the target sequences.

Template (GD): The target DNA sequence that is the amplification objective of a PCR. GD is composed of four letters (A, T, C, and G) that represent each one of the nucleotides, which are listed in order of their appearance in DNA. The infinite specificity of the targets proceeds from variations of the length and order of letters in a GD. In PCR experiments, the GD length is <5,000 bp. GD' is the complementary strand for GD, where T is the complementary letter for each A, C is complementary for each G, and *vice versa*. Thus, for a GD (GD = ATTAAGGCCATCG...), the GD'=TAATTCGGTAGC... [28].

Sense primer of GD or B_f = {bi | i is the index of GD between F_s and F_e}, where F_s and F_e are respectively the first and last bases of the sense primer B_f in GD. The anti-sense primer of GD or B_r = {bi | i is the index of GD between R_s and R_e}, where R_s is the first letter and R_e is the last letter of B_r in GD. The individuals or possible solutions of the minimization problem are denoted as a primer pair, which is presented as a P_i vector [28]:

$$P_i = (F_s, \alpha, \beta, \gamma) \quad (1)$$

where α is the number of bases of B_f; β is the number of bases between B_f and B_r; and γ is the number of bases in the primer B_r. The independent variables are defined by (2), (3), and (4), and these equations represent the relative position of each primer [28]:

$$\alpha = (F_e - F_s) \quad (2)$$

$$\beta = (R_s - F_e) \quad (3)$$

$$\gamma = (R_e - R_s) \quad (4)$$

Primer length (tam): In a P_i primer of GD, the length |tam| is the summation of the number of bases in the primer and is obtained from Eq. (5) [28]:

$$|\text{tam}| = \#G + \#C + \#A + \#T \quad (5)$$

Melting temperature (Tm): The Tm estimates the DNA-DNA duplex stability and can be calculated in accordance with the thermodynamic nearest-neighbor (NN) model of the fusion process with unified ΔH and ΔS [28], as given by Eq. (6):

$$Tm(^{\circ}C) = \frac{\Delta H}{\Delta S + R \ln(C)} + 12 \log_{10}[\text{Na}^+] - 273.15 \quad (6)$$

where C is the primer concentration in empirically calculated

molarity, R is the molar constant of the gases (1.987 cal/K mol), ΔH (kcal/mol) is the enthalpy change, and ΔS is the entropy change for the helix formation. Here, ΔH and ΔS are calculated using the NN model under standardized conditions (250 mMol/l NaCl, pH 7.0) [3]. The stabilizing effect of the salt (sodium) on the duplex is given by

$$\Delta S^{\circ} (\text{Salt correction}) = \Delta S(1M \text{ NaCl}) + 0.368 * N * \ln([\text{Na}^+]) \quad (7)$$

where N is the number of nucleotide pairs in the primer (primer length -1) expressed in bp, and $[\text{Na}^+]$ is the sodium equivalent in mM calculated by Eq. (8):

$$[\text{Na}^+] = \text{Na}^+ \text{Concentration} * N * \ln([\text{Na}^+]) \quad (8)$$

Annealing temperature (Ta): The fusion or annealing temperature of the primer on GD. The optimal Ta is calculated according to Rychlik's formula (Eq. 9) [28]:

$$Ta = 0.3 * Tm (\text{primer}) + 0.7 Tm (\text{amplicon}) - 14.9 \quad (9)$$

where Tm is the melting temperature.

2) Select the first primer pair: first optimization. The second step in the Multiplex-PCR primer design process by Mult-PSOS is the choice of the first primer pair. This procedure follows the classical minimization process described in Refs. [9] and [28] with some modifications. One of the main differences between these works and the current study is that the optimization technique employed in this study does not require the definition of any heuristic parameter (mutation probability, crossover rate, and selection rules in genetic algorithms, or cognitive parameters and inertia weight in PSO) in order to obtain a solution.

The original fitness function [28] and its restrictions described in Ref. [26] are defined as follows:

$$\text{Fitness}(P_i) = \text{tam}(P_i) + 3\text{tamd}(P_i) + 3\text{Tmd}(P_i) + 3\text{GC}_p(P_i) + 3(\text{GCc}(\text{Bf}) + \text{GCc}(\text{Br})) + 50\text{Uni}(P_i) + 10\text{Sc}(P_i) + 10\text{PC}(P_i) + \text{Rt}(P_i) \quad (I)$$

where (Pt) is the primer pair; tam(Pt) is the primer length; tamd(P_i) is the length difference between the two primers of a pair; Tmd(P_i) is the Tm difference into a primer pair; GC_p(P_i) is the GC%; GCc(Bf) + GCc(Br) is the G, C, CG, or GC presence in the 3' end (Bf) or (Br) primer; Uni(Pt) is the primer specificity; Sc(Pt) indicates the presence of self-complementary sequences; PC(P_i) indicates the presence of complementary structures within the pair; and Rt(Pt) is the restriction site check function.

In order to improve the primer's sensitivity, specificity, and efficiency in Multiplex-PCR for infectious disease diagnosis and through analyzing previous studies [1, 9, 22, 31], it was possible for us to define new terms and different weights for the parameters described in Eq. (I) [28].

The modified fitness function is presented in Eq. (A):

$$\begin{aligned} \text{Fitness}(P_i) = & \text{tam}(P_i) + \text{tamd}(P_i) + 3\text{Ta}(Pt) + 10\text{Tad}(Pt) + 3\text{Tm}(Pt) \\ & + 10\text{Tmd}(P_i) + \text{GC}_p(P_i) + 3\text{Term}(Pt) + 3\text{GCc}(Pt) \\ & + 50\text{Uni}(Pt) + 10\text{Sc}(Pt) + \text{PC}(P_i) + \text{PCh}(P_i) + \text{Rep}(Pt) \\ & + \text{la}(Pt) \end{aligned} \quad (\text{A})$$

where $\text{Tm}(Pt)$ is the melting temperature; $\text{Ta}(Pt)$ is the annealing temperature of a primer in Pt ; $\text{Tad}(Pt)$ is the $\text{Ta}(Pt)$ difference of the members in a primer pair; $\text{Term}(Pt)$ is the 3' end stability related to the ΔG value (a divergent definition of the $\text{Term}(Pt)$ presented in Eq. (I)); $\text{PCh}(P_i)$ is the number of internal loops with a $\Delta G \geq -6\text{kcal/mol}$; and $\text{Rep}(Pt)$ is the false priming prediction due to base repetition in tandem. Finally, $\text{la}(Pt)$ is the amplicon length. These factors take into account the kinetic of the primer-template duplex. Different weights from Ref. [26] were assumed for $\text{tam}(Pt)$, $\text{tamd}(Pt)$, $\text{GC}_p(Pt)$, and $\text{PC}(Pt)$. The weights (3, 1, 3, 10, and 50) in Eq. (A) were applied considering the relevance of the primer restrictions. These values are set based on realistic conditions and on the studies mentioned above [1, 9, 28].

After inserting the target sequence (GD), 10 pairs of primers were defined as the initial population (see Eqs. (1)–(4)). Within this population, the Mult-PSOS chooses, at each iteration, the global best (the best primer pair in the population; see Eqs. (a) and (b)) until the process completes after 200 iterations (the stopping criterion). As a result of the minimization process, a first optimum primer pair is selected.

Restrictions for the first optimization (first primer design in a PCR)

At the end of this stage, the last primer pair and its optimum parameters (tam , Ta , and Tm) are determined (again, the optimization process was carried out employing the Mult-PSOS algorithm). The restrictions applied in the first optimization are defined below:

- Primer length (tam):** The absolute value of tam indicated by Eq. (5) is defined as follows: if $18 \leq \text{tam}(Pt) \leq 26 \rightarrow \text{tam}(Pt) = 0$ otherwise $\text{tam}(Pt) = 1$ [28].
- tamd :** The difference between $\text{Tam}(\text{Bf})$ and $\text{Tam}(\text{Br})$ is defined as $\text{tamd}(Pt)$. The absolute value of tamd is calculated using Eq. (10) and represents the difference in length between Bf and Br [28].

$$\text{ABS}[\text{tamd}(\text{Bf}) - \text{tamd}(\text{Br})] \leq 2 \quad (10)$$

The value of tamd between Bf and Br is defined as $\text{tamd}(Pt) = 0$, if the absolute value indicated in Eq. (10) $\text{ABS} \leq 2$, otherwise $(\text{tamd})Pt = 1$.

- Tmd :** The Tm difference between Bf and Br should be $\leq 5^\circ\text{C}$. Tmd is defined by Eq. (11) [28]:

$$\text{ABS}[\text{Tm}(\text{Bf}) - \text{Tm}(\text{Br})] \leq 5 \quad (11)$$

Tmd between Bf and Br is defined as $\text{Tmd}(Pt) = 0$, if the absolute value indicated in Eq. (12) is ≤ 5 , otherwise $\text{Tmd}(Pt) = 1$ [28].

- $\text{GC content (GC}_p)$:** In a primer, 40% to 60% of the bases should be G or C. The GC_p is given by Eq. (12) [28]:

$$|\text{GC}_p|(Pt) = (\#G + \#C) * 100/\text{tam} \quad (12)$$

In the check function of GC content in a primer, GC_p is denoted as $|\text{GC}_p|(Pt)$ and has a null value $|\text{GC}_p|(Pt) = 0$, if $40\% \leq \text{GC}(\text{Bf})$ and $\text{GC}(\text{Br}) \leq 60\%$, otherwise, $|\text{GC}_p|(Pt) = 1$.

- GC clamp (GCc) :** The 3' end of the primer Pt is denoted as $\text{GCc}(Pt)$, and in the check function $\text{GCc}(Pt)$ is defined as $\text{GCc}(Pt) = 0$, if the 3' end is G or C, otherwise $\text{GCc}(Pt) = 1$. The 3' end should be 0, where $\text{GCc}(\text{Bf}) = 0$, $\text{GCc}(\text{Br}) = 0$, $\text{GCc}(Pt) = \text{GCc}(\text{Bf}) + \text{GCc}(\text{Br})$.
- Specificity ($\text{Uni}(Pt)$):** The primer Pt should be coupled in only one place inside GD . This property is denoted by $\text{Uni}(Pt) = 0$, if Pt is coupled once in GD , otherwise $\text{Uni}(Pt) = 1$. $\text{Uni}(Pt)$ should be 0 for any of the primers $\text{Uni}(\text{Bf}) = 0$, $\text{Uni}(\text{Br}) = 0$, $\text{Uni}(Pt) = \text{Uni}(\text{Bf}) + \text{Uni}(\text{Br})$ [28].
- Secondary structures in the primer:** These structures can be produced by intermolecular interactions (bugle loops and internal loops) or intramolecular interactions (hairpin loops) of the primer. Secondary structures relevant to PCR performance are those that are stable over the primer annealing temperature. The stability of these loops is represented by the ΔG value (Eq. 13) [28]:

$$\Delta G = \Delta H - \Delta S \quad (13)$$

where ΔH and ΔS are the enthalpy and entropy changes of the primer, respectively. Large negative values indicate a stable structure or undesirable loops (such as loops proximal to the 3' end).

- Self-complementary sequences or hairpin loops ($\text{Sc}(Pt)$):** In the primer design, these kinds of sequences should be avoided. Thus, self-complementary sequences should be null ($\text{Sc}(Pt) = 0$) if there is no self-complementarity in Bf and Br , otherwise $\text{Sc}(Pt) = 1$ [28].
- Heterodimers ($\text{PC}(Pt)$):** Heterodimers with a $\Delta G \geq -6\text{kcal/mol}$ are tolerable. The heterodimer presence is given by $\text{PC}(Pt)$, where $\text{PC}(Pt) = 0$ if there is no complementarity between primers Bf or Br ; otherwise $\text{PC}(Pt) = 1$ [28].

Additionally, the following restrictions are taken into account in this work:

- Hybridization temperature (Ta):** Ta is defined as follows: if $56^\circ\text{C} \leq \text{Ta}(Pt) \leq 64^\circ\text{C} \rightarrow \text{Ta}(Pt) = 0$, otherwise $\text{Ta}(Pt) = 1$ [28].
 - Tad :** The difference between $\text{Ta}(\text{Bf})$ and $\text{Ta}(\text{Br})$ is defined as $\text{Tad}(Pt)$ and its absolute value is given by Eq. (14) [28].
- $$\text{ABS}[\text{Ta}(\text{Bf}) - \text{Ta}(\text{Br})] \leq 5 \quad (14)$$
- Tad between Bf and Br is defined as $\text{Tad}(Pt) = 0$ if the absolute value indicated in Eq. (11) is ≤ 5 , otherwise $\text{Tad}(Pt) = 1$.
- Tm :** $\text{Tm}(Pt)$ of Bf and Br is defined as $\text{Tm}(Pt) = 0$ if $56^\circ\text{C} \leq \text{Tm}(Pt) \leq 64^\circ\text{C}$, otherwise $\text{Tm}(Pt) = 1$ [23].
 - 3' end ΔG (Term):** The 3' stability is given by the ΔG value or $\text{rm}(Pt)$. Both low and high negative values of the 3' end

primer ΔG can promote mistakes by instability or unspecific annealing. Then, if $4.5 \leq \Delta G \leq 9.9$, $Term(Pt) = 0$, otherwise $Term(Pt) = 1$.

l. *Amplicon length la(Pt)*: Is the length of the PCR product and is calculated by Eq. (15):

$$la(Pt) = (\alpha + \beta + \gamma) + 1 \tag{15}$$

where α is the number of bases of Bf, calculated by Eq. (2); β is the number of bases between Bf and Br (Eq. 3); γ is the number of bases in the primer Br (Eq. (4)). For the Multiplex-PCR, the $la(Pt)$ should be between 200 and 800 bp [26]. In the check function, $la(Pt) = 0$ if $200 \leq la(Pt) \leq 800$, otherwise $la(Pt) = 1$.

m. *Homodimers (PCh(Pt))*: Bugle loops or internal loops with a $\Delta G \geq -6$ kcal/mol are tolerable. The homodimer presence is given by PCh(Pt), where PCh(Pt) = 0 if there is no complementarity between the same kind of primers of Bf or Br; otherwise PCh(Pt) = 1.

n. *Repetitions (Rep(Pt))*: In a primer design, tandem repetitions of four or more of the same kind of bases should be avoided. Rep(Pt) = 0 if there is no repetition, otherwise Rep(Pt) = 1 [28].

o. *Cross homology*: In a primer sequence, regions with unspecific homology with other genes (that are not templates) present in the reaction should be avoided. For that reason, the primer sequence designed should be run by nucleotide BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_SPEC=WGS&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch), with the appropriate organism or database (according to the specific case) for identification of homology regions in each primer [4].

3) Multiplex-PCR optimization process: second optimization.

In the second optimization step, the *fitness* given by Eq. (B) is minimized (for the second, third, and *n*-th primers) by applying the restrictions defined in the first optimization step and using, as always, 10 pairs of primers and 200 iterations. At this point, the first primer pair (obtained in the first optimization step) and its optimum parameters (the tam , Ta , and Tm values) are now used as restrictions for all the other primers (see Eqs. (5)–(18)).

$$Fitness(Mult) = \sum Fitness(P_i) + tamM(P_i) + 10 * TamM + 10 * TmdM(P_i) + 50 * laM(Pt) \tag{B}$$

In this manner, the process continues until the last primer pair is obtained. This strategy guarantees the rapid mathematical convergence of the method. Equation (B) represents the Multiplex-PCR primer design strategy proposed in this work. The Multi-PSOS was coded in MATLAB (Fig. 2) and run on an Intel Core 2 Duo – 1.33 GHz processor with 2.00 GB of RAM. The results obtained with the proposed algorithm were compared with the

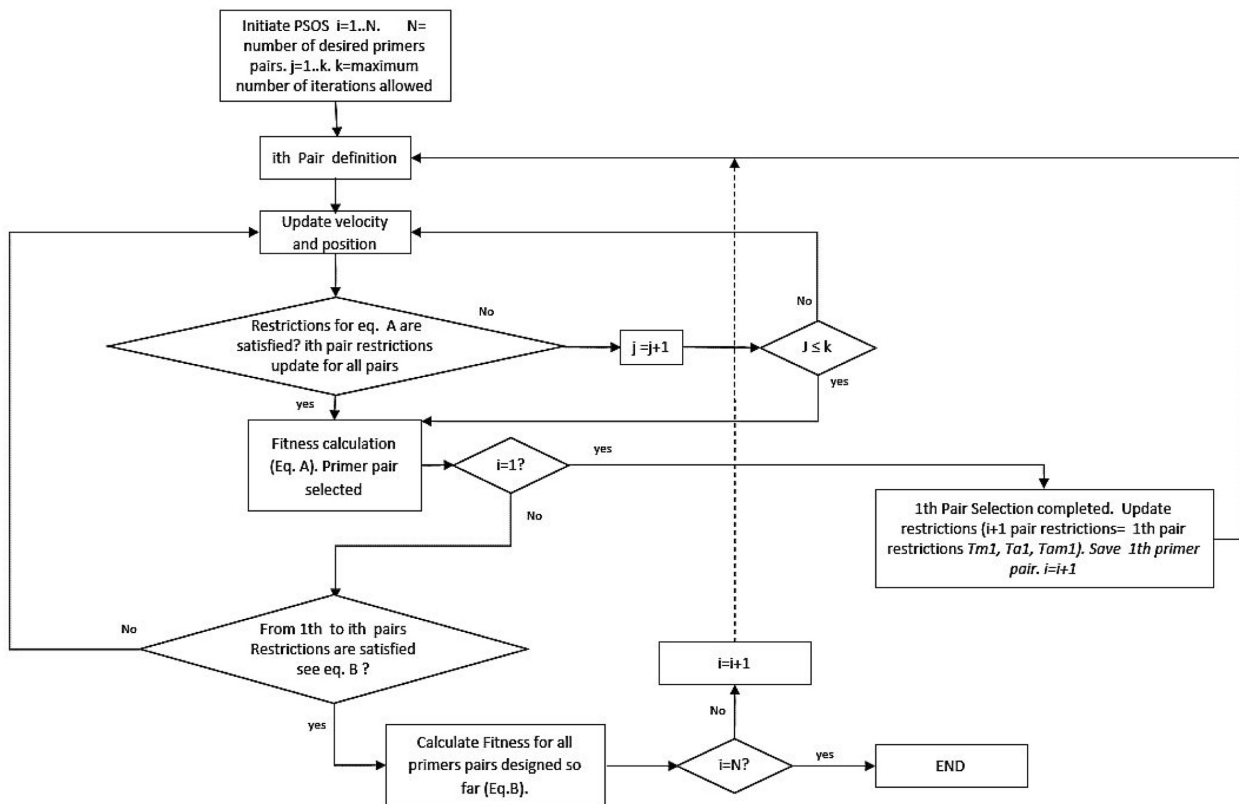


Fig. 2. Basic flow diagram of the Multi-PSOS algorithm.

primers described by Ref. [28] and the oligonucleotides obtained using the Primer-BLAST tool (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>).

Restrictions for the primer design in a Multiplex-PCR (second optimization step)

For a Multiplex-PCR where multiple primer pairs must be selected, the first pair of oligonucleotides is selected by minimizing Eq. (A). Based on the parameters of the selected primers, the t_{am} , T_a , and T_m values are fixed as initial values for all the primers. In a second optimization, the fitness given by Eq. (B) is recalculated with the restrictions that defined the identification of the second primer pair and so on. These restrictions are defined as follows:

a. $ladM(P_t)$: The size difference among all the amplicons replicated in a Multiplex-PCR. The minimum $ladM(P_t)$ should be 50 bp. The $la(P_t)$ of the first primer pair is defined by Eq. (16). The second amplicon length ($la(P_t)$) should meet the length restriction of the $la(P_t)_1$ and, in addition, $la(P_t)_2$ should be within the following interval: $la(P_t)_2 = la(P_t)_1 \pm 50$ bp and $la(P_t)_2 \neq la(P_t)_1$. For the third primer pair selection, the $la(P_t)$ restriction is given by $la(P_t)_{n+1} = la(P_t)_n \pm 50$ bp and $la(P_t)_{n+1} \neq la(P_t)_n = la(P_t)_{n+1} = 0$, otherwise $la(P_t)_{n+1} = 1$.

b. $tamdM(P_t)$: The primer length difference $tam(P_t)$ among all the primers in the Multiplex-PCR ($Bf_{1,2,...,n}$ and $Br_{1,2,...,n}$). For the selection process, the smallest primer (or the one with lowest value of la) in a Multiplex-PCR is fixed in each run and all the selected primers, in the next run, should have a number of bases ≤ 2 . The absolute value of $tamdM(P_t)$ is calculated by Eq. (16):

$$ABS[tam(Bf_{1,2,...,n}) \text{ or } tam(Br_{1,2,...,n}) - tam_{lowest}] \leq 2 \quad (16)$$

The difference in the number of bases ($tamdM(P_t)$) among all primers in a Multiplex-PCR ($Bf_{1,2,...,n}$ or $Br_{1,2,...,n}$) is defined as $tamdM(P_t) = 0$ if the absolute value indicated in Eq. (18) is ≤ 2 , otherwise $tamdM(P_t) = 1$.

c. $TadM(P_t)$: The absolute value of $Ta(P_t)$, the annealing temperature difference among all the primers in a Multiplex-PCR ($Bf_{1,2,...,n}$ and $Br_{1,2,...,n}$). For the selection process, the primer with the lowest $Ta(P_t)$ (in a Multiplex-PCR) is fixed in each run and all the selected primers, in the next run, should have a $Ta(P_t) \leq 6$. The absolute value of $tadM(P_t)$ is calculated

according to Eq. (17):

$$ABS[Ta(Bf_{1,2,...,n} \text{ or } Br_{1,2,...,n}) - Ta_{lowest}] \leq 6 \quad (17)$$

The difference in the number of bases ($TadM(P_t)$) among $Bf_{1,2,...,n}$ or $Br_{1,2,...,n}$ is defined as $TadM(P_t) = 0$ if the absolute value indicated in Eq. (17) is ≤ 6 , otherwise $TadM(P_t) = 1$.

d. $TmdM(P_t)$: The absolute value of $Tm(P_t)$, the primer length difference among all the primers in a Multiplex-PCR ($Bf_{1,2,...,n}$ and $Br_{1,2,...,n}$). For the selection process, the primer with the lowest Tm (in a Multiplex-PCR) is fixed in each run and all the selected primers, in the next run, should have a $Tm(P_t) \leq 6$. The absolute value of $TmdM(P_t)$ is calculated using Eq. (18):

$$ABS[Tm(Bf_{1,2,...,n} \text{ or } Br_{1,2,...,n}) - Tm_{lowest}] \leq 6 \quad (18)$$

The difference in the number of bases ($TmdM(P_t)$) among $Bf_{1,2,...,n}$ or $Br_{1,2,...,n}$ is defined as $TmdM(P_t) = 0$ if the absolute value indicated in Eq. (18) is ≤ 6 , otherwise $TmdM(P_t) = 1$.

e. *Checking function for Multiplex-PCR primer design*: The checking process is accelerated by using a matrix that registers the position and the specific value of *fitness* of each primer pair. The value of the objective function (*fitness*) is given by Eq. (B) and is calculated using the restriction design described above in Eqs. (16)–(18).

4) Stopping criterion. In a Mult-PSOS process, each individual of the particle swarm is defined through the vector given by Eq. (1) and its fitness is evaluated by Eqs. (A) in the first optimization step or (B) in the second step. The optimization process described by Eqs. (A) and (B), with its restrictions, continues until the defined stopping criterion is met. In this case, we use 200 iterations for the PSO and 200 iterations for the Simplex algorithm.

Bacteria and Gene Sequences

DNA sequences from genes of uropathogenic bacteria consigned in the GenBank (<http://www.ncbi.nlm.nih.gov/genbank>) database were used as templates for *in silico* primer design experiments by PSOS and for *in vitro* PCR assays (Table 1). The selection of the bacteria and the genes was made according to Ref. [20]: *E. coli FimH* gene (GenBank access AJ225176); *K. pneumoniae FimK* gene (GenBank access EU315065.1); *P. aeruginosa ETA* gene

Table 1. ATCC bacterial lineages and genes used as a model for the Multiplex-PCR primer design by Mult-PSOS.

Bacteria	ATCC code	Gene (GD)	Gene function	GenBank Access. No.	Ref.
<i>E. coli</i>	25922	<i>FimH</i>	Type I fimbriae	AJ225176	[14]
<i>K. pneumoniae</i>	13883	<i>FimK</i>	Type I fimbriae	EU315065.1	[20]
<i>P. aeruginosa</i>	27853	<i>ETA</i>	Exotoxin A	K01397.1	[20]
<i>P. mirabilis</i>	12453	<i>ZapA</i>	Metalloprotease	AF064762.1	[25]
<i>P. mirabilis</i>	12453	<i>UreC</i>	Urease	AM942759	[25]
<i>S. saprophyticus</i>	15305	<i>16S</i>	16 S rRNA	NR_074999.1	[17]
<i>E. faecalis</i>	29212	<i>16S</i>	16 S rRNA	NR_074637.1	[17]

(GenBank access K01397.1); and *P. mirabilis* UreC gene (GenBank access AM942759). For the Multiplex-PCR design in this study, two uropathogenic bacteria were added: *S. saprophyticus* (16S gene, GenBank access NR_074999.1) and *E. faecalis* (16S gene, GenBank access NR_074637.1) [10, 17].

Bacterial Growth and Preservation

For the *in vitro* experiments, the gram-negative bacteria (*E. coli*, *K. pneumoniae*, *P. aeruginosa*, and *P. mirabilis*) were grown in Luria-Bertani broth and agar. Muller-Hinton broth and agar were used for the growth of the Gram-positive bacteria (*S. saprophyticus* and *E. faecalis*). Microorganisms in mid-log phase ($OD_{620} = 0.4$) were preserved in 10% glycerol at -80°C pending further use.

DNA Extraction

For DNA extraction, all the microorganisms (except *E. coli*) in mid-log phase growth were pelleted and washed two times with sterile $1\times$ PBS. The washed and pelleted cells were resuspended in 2 volumes of PrepMan Ultra (PMU) reagent (Applied Biosystems). Then, cells in suspension were disrupted by the freeze-thaw technique (3 min in a liquid nitrogen bath followed by 10 min at 95°C). The DNA extract was obtained by ultracentrifugation at 14,000 rpm for 2 min and preserved at -20°C with RNase A ($10\ \mu\text{g}/\text{ml}$) until the PCR experiments. DNA extraction for *E. coli* was obtained directly by the colony boiling method. The genomic DNA quality was confirmed on a 1% agarose gel stained with SYBR Green and quantified using a NanoDrop ND-1000 UV-Vis spectrophotometer (Thermo Scientific).

Wet Experiments of Multiplex-PCR

The optimal PCR primers designed were synthesized and used for the Multiplex-PCR assays. Table 1 summarizes the sequence of the selected primer pairs (obtained by minimizing the fitness functions given by Eqs. (A) and (B) subject to the restrictions in Eqs. (1) to (19)) for the amplification of the specific gene for each uropathogen. For the Multiplex-PCRs, we used an amplification mixture containing $4\ \mu\text{l}$ of template (50–200 ng DNA), 0.4–0.6 mM of each primer, 1.5 U of GoTaq Flexi DNA Polymerase (Promega), and 1.5 mM MgCl_2 for a final volume of $50\ \mu\text{l}$. Cycling reactions were performed under the following conditions: one cycle of 4 min at 94°C , followed by 35 cycles of 60 sec at 94°C , 50 sec at 53°C , and 70 sec at 72°C . Finally, one post-amplification step of 10 min at 72°C was carried out. The amplicon quality and size were verified on 2% agarose gels and revealed by SYBR Gold stain (Molecular Probe, Invitrogen Life Technologies) with 1 Kb and 100 pb DNA Ladder (Thermo Scientific) as a molecular weight marker. The amplicons were ligated to the cloning vector pCR 2.1 TOPO (Invitrogen Life Technologies). The sequence of the insert was confirmed by DNA sequencing and a search by homology was then performed using the BLASTn algorithm (<http://www.ncbi.nlm.nih.gov/BLAST>) [4]. For the verification of the primer specificity, extracted DNA from *Serratia marcescens*, *Salmonella* spp., and *Staphylococcus aureus* were used as negative

controls in the Multiplex-PCR.

Results

Primer Design for Multiplex-PCR

In this study, a set of Multiplex-PCR primers for uropathogenic bacteria detection were designed using a Mult-PSOS tool. In Tables 2 and 3, the sequences and some of the theoretical primer design parameters calculated by Gene Runner 3.01 are presented for the group of oligonucleotides. All the aforementioned design restrictions were evaluated for the 12 primers selected by PSOS. The conditions derived were evaluated using Eqs. (A), (B), and (1)–(18). The final fitness checks for the PCR and Multiplex-PCR were made using Eqs. (A) and (B), respectively. The fitness values thus obtained are presented in Table 4.

In Vitro Multiplex-PCR Assays

The primer sequences predicted by the hybrid algorithm were synthesized and used for *in vitro* replication assays. The Ta(Pt) and la(Pt) parameters calculated *in silico* were used for calibration of the cycle conditions (number of cycles, temperature, and time for each of the PCR phases). The Multiplex-PCR was designed for the amplification of seven genes from six pathogens highly prevalent in the causation of UTIs. The electrophoretic analysis on the agarose gel (2%) showed a good resolution between bands with amplicons ranging from 235 to 740 bp, and a minimum difference of 50 bp among them. The electrophoretic profile of the Multiplex-PCR showed an adequate performance of the amplification. Densitometry analysis of the amplicon bands revealed a DNA concentration ranging from 67 to 128 ng/ μl for the *E. coli* *FimH* gene and *E. faecalis* 16S gene, respectively (Fig. 3).

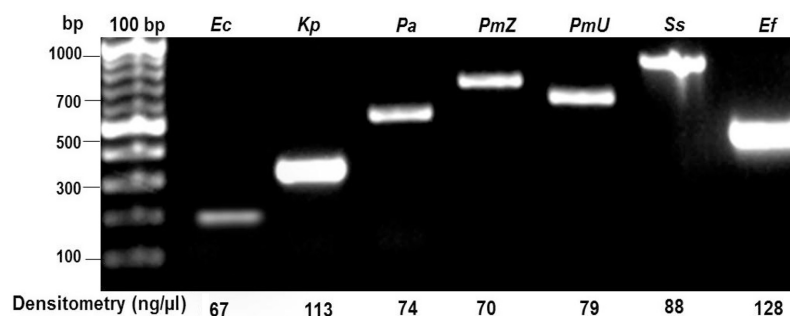
Benchmarking

For the benchmarking analysis, the parameters and fitness values (from Eqs. (A) and (B)) obtained from the UTI Mult-PSOS primers were compared with two groups of oligonucleotides, one designed using the online tool primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) and the other described previously by Padmavathy *et al.* [20]. Five genes from four uropathogenic bacteria (*E. coli* (*FimH*), *K. pneumoniae* (*FimK*), *P. aeruginosa* (*ETA*), and *P. mirabilis* (*ZapA* and *UreC*)) were used as *GD*. For the primer-BLAST estimation, the PCR product size was adjusted with 100 bp and 800 bp as the minimum and maximum lengths, respectively. The first primer pair generated for each DNA template (gene) by primer-BLAST

Table 2. General parameters of uropathogenic Multiplex-PCR primers.

Bacteria	Gene (GD)	Primers (5'-3')	Primer position in GD	Amplicon (bases)	Length la (bases)	Ta (°C)	Tm (°C)*	GCp (%)	3'End ΔG kCal/mol	Ref.
<i>E. c</i>	<i>FimH</i>	F CTTTATTGACCCGGTGAGC	1721-1741	235	21	51.8	59.4	52.4	-7.6	This Study
		R CGGTAAGAGGAATCGGCACTG	1935-1955							
<i>K. p</i>	<i>FimK</i>	F GGGTGAAAAAGCTGGTGAATG	548-569	315	22	52.8	60.4	50.0	-6.9	
		R TAATATTCTGCACGGCTCCTG	841-862							
<i>P. a</i>	<i>ETA</i>	F CCTCATCTTCACCCATCACAG	713-734	505	22	50.6	58.2	54.5	-4.8	
		R GATCGGCGACATGTGGCTGAG	1196-1217							
<i>P. m</i>	<i>ZapA</i>	F GCGATCCCACGTTAGTTGTC	5951-5971	571	21	51.0	58.6	57.1	-4.6	
		R AGTTTGCCAGCTGATACCGG	6501-6521							
	<i>UreC</i>	F ATTATGGAAGGCGCACGAGA	1413-1432	649	20	49.8	57.4	50.0	-6.6	
		R ACCAAATTTGACCTCTTCGCC	2041-2061							
<i>S. s</i>	<i>16S</i>	F TACATGCAAGTCGAGCGAACAG	52-73	741	22	50.6	58.2	50.0	-6.9	
		R CTGTTTGATCCCCACGCTTTC	772-792							
<i>E. f</i>	<i>16S</i>	F TTCGGGTGTCGTTGATGGATG	221-241	440	21	52.6	60.2	52.4	-6.5	
		R CTCCAGTTTCCAATGACCC	641-660							
<i>E. c</i>	<i>FimH</i>	F TCGAGAACGGATAAGCCGTGG	1691-1711	508	21	54.1	61.7	57.1	-11.1	[20]
		R GCAGTCACCTGCCCTCCGGTA	2176-2196							
<i>K. p</i>	<i>FimK</i>	F TGCTCTATCAGGTGAGTCAT	275-294	746	20	38.1	45.7	45	-4.5	
		R AAAATCGATAGTTTCAGCAT	1011-1020							
<i>P. a</i>	<i>ETA</i>	F GCCTCCGAACATCAAGGTGT	1147-1166	207	20	43.7	51.3	61.1	-6.1	
		R CCATGACCACGCTGACC	1333-1349							
<i>P. m</i>	<i>ZapA</i>	F ACCGCAGGAAAACATATAGCCC	2981-3002	540	22	52.1	59.7	50.0	-8.3	
		R GCGACTATCTCCGCATAATCA	3499-3520							
	<i>UreC</i>	F GTTATTCGTGATGGTATGGG	2071-2090	317	20	41.7	49.3	45.0	-6.8	
		R ATAAAGGTGGTTACGCCAGA	2368-2387							

*Thermodynamic Tm, calculated with 250 mMol of salt concentration (Na⁺). F: forward; R: reverse. Ec: *E. coli*. Kp: *K. pneumoniae*. Pa: *P. aeruginosa*. Pm: *P. mirabilis*. Ss: *S. saprophyticus*. Ef: *E. faecalis*.

**Fig. 3.** Agarose gel electrophoresis (2%) for uropathogenic Multiplex-PCR products amplified with the primer sequences designed by Mult-PSOS.

1) 1 kb molecular weight ladder (Thermo Scientific). Ec: *E. coli* (*FimH*), 235 bp. Kp: *K. pneumoniae* (*FimK*), 316 bp. Pa: *P. aeruginosa* (*ETA*), 505 bp. PmZ: *P. mirabilis* (*ZapA*), 571 bp. PmU: *P. mirabilis* (*UreC*), 649 bp. Ss: *S. saprophyticus* (*16S*), 741 bp. Ef: *E. faecalis* (*16S*), 440 bp.

Table 3. Structural parameters analysis of uropathogenic Multiplex-PCR primers.

Bacteria	Gene	Primer	Secondary structures												Ref.			
			Dimers			Base runs	Hairpin loops			Bulge loops			Internal loops					
			#	ΔG kcal/Mol	Tm (°C)		#	ΔG kcal/Mol	Tm (°C)	#	ΔG kcal/Mol	Tm (°C)	#	ΔG kcal/Mol		Tm (°C)		
<i>E. c</i>	<i>FimH</i>	F	3	-3.0	-74	-	2	-3.6	+75	-	-	-	-	-	-	[20]		
				-3.0	-74			-0.8	+43									
				-1.8	-97													
		R	5	-4.8	-57	-	3	-0.7	+45	1	-4.5	-73.1	1	-4.9	-37			
				-1.6	-93			-0.7	+38									
			-1.6	-93			-0.3	+32										
			+0.6	-128														
			+0.6	-128														
	<i>K. p</i>	<i>FimK</i>	F	6	+1.5	-164	-	1	+0.9	0.0	1	+1.6	-107.2	3	+0.8		-103	
					+1.5	-164			+2.2	-111								
+1.5					-164	+2.6			-111									
+1.5					-164													
			+1.8	-154														
			+1.8	-154														
		R	4	-4.8	-48.	-	2	+0.3	+20.5	2	-3.4	-29.8	1	+1.8	-76			
			+2.2	-128	+0.3			+20.5	-2.4		-42.0							
		+2.2	-128															
		+2.2	-128															
<i>P. a</i>	<i>ETA</i>	F	4	+0.3	-122	-	2	-0.6	+36.9	1	-1.0	-86.1	1	-2.2	-68			
				+0.3	-122			+1.2	0.0									
				+1.8	-159													
				+1.8	-159													
<i>P. m</i>	<i>ZapA</i>	R	1	-0.3	-11	-	-	-	-	-	-	-	-	-	-			
		F	2	+2.1	-110	-	-	-	-	-	-	-	-	-	-			
					+2.7	-121												
		R	4	-1.7	-97	-	2	-2.5	+71.6	1	-0.9	-69.0	2	-6.5	-40			
					-1.7	-97			+1.7	0.0				+4.6	-102			
					+3.6	-150												
					+3.6	-150												
		<i>UreC</i>	F	-	-	-	-	-	-	-	-	-	-	-	-			
		R	4	0.0	-129	-	2	-0.6	+40.8	1	-0.8	-84.6	2	-2.1	-72			
				0.0	-129			+1.5	0.0				+3.9	-96				
				+3.2	-146													
				+3.2	-146													

and all the primer sequences described by Ref. [20] were subjected to oligonucleotide analysis by Gene Runner 3.01. The good performance of the Mult-PSOS can be verified by the comparison of the fitness values calculated for the three groups of primers. The results of the check functions from

Eqs. (A) and (B) are shown in Tables 3 and 4.

Discussion

In Multiplex-PCR assays, a vast number of parameters

Table 3. Continued.

Bacteria	Gene	Primer	Secondary structures												Ref.		
			Dimers			Base runs	Hairpin loops			Bulge loops			Internal loops				
			#	ΔG kcal/Mol	Tm (°C)		#	ΔG kcal/Mol	Tm (°C)	#	ΔG kcal/Mol	Tm (°C)	#	ΔG kcal/Mol		Tm (°C)	
<i>E. c</i>	<i>FimH</i>	F	1	-4.8	-57	-	-	-	-	-	-	-	-	-	-	-	This study
		R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
<i>K. p</i>	<i>FimK</i>	F	1	-1.3	-89	1	-	-	-	-	-	-	-	-	-	-	
		R	3	-4.3	-61	-	-	-	-	-	-	-	-	-	-	-	
				-1.9	-97												
				-1.7	-55												
<i>P. a</i>	<i>ETA</i>	F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
		R	4	-2.9	-67	-	-	-	-	-	-	-	1	-0.8	-68		
				-0.2	-125												
				-0.2	-125												
				+0.3	-118												
<i>P. m</i>	<i>ZapA</i>	F	2	-1.2	-94	-	-	-	-	-	-	-	-	-	-	-	
				-0.3	-118												
		R	2	-5.1	-49	-	-	-	-	-	-	-	-	-	-	-	
				-4.8	-57												
	<i>UreC</i>	F	1	-4.8	-56	-	-	-	-	-	-	-	-	-	-	-	
		R	1	-7.9	-20	1	-	-	-	-	-	-	-	-	-	-	
<i>S. s</i>	<i>16S</i>	F	5	-1.9	-97	-	-	-	-	-	-	-	1	-1.2	-68		
				-1.8	-97												
				-0.3	-113												
				-0.2	-125												
				-0.2	-125												
		R	1	+0.3	-118	1	-	-	-	-	-	-	-	-	-	-	
<i>E. f</i>	<i>16S</i>	F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
		R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

F: forward; R: reverse. Ec: *E. coli*. Kp: *K. pneumoniae*. Pa: *P. aeruginosa*. Pm: *P. mirabilis*. Ss: *S. saprophyticus*. Ef: *E. faecalis*.

play a critical role in the efficiency of the DNA target amplification. These parameters include (i) optimal nucleotides and enzyme concentration in the reaction; (ii) quality and amount of extracted DNA template; (iii) number of PCR cycles; (iv) quantity of primers (inverse to amplicon size); and (v) salt concentration [26]. Nevertheless, optimal primer design is the most relevant factor in Multiplex-PCR performance.

Currently, software is available for the solution of the Multiplex-PCR primer design problem, but, in general,

these options are non-free closed algorithms that do not take into account all the necessary restrictions for the adequate prediction of these sequences. Previously, a Multiplex-PCR primer design strategy based on a genetic algorithm for tuberculosis diagnosis has been described [29]. However, in that work, the AG control parameters known as mutation probability (mp) and crossover probability (cp) must be defined by the user. The designation of these initial parameters requires *a priori* knowledge of AG, which constitutes a serious inconvenience for the

Table 4. Checking function results and fitness comparison for three sets of uropathogenic Multiplex-PCR primers.

	Primer	Tam	Tamd	Ta	Tad	Tm	Tmd	CGp	³ End dG	Term	Uni	NoSc	Sc	PCh	Rep	la	Fitness (Pt)	TamdM	# TamdM	TmdM	# TadM	TadM	laM	Fitness Mult	Ref.	
<i>E. c.</i>	<i>FimH F</i>	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	13	1	7	1	7	1	0	293	[20]	
	<i>FimH R</i>	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	4									
<i>K. p.</i>	<i>FimK F</i>	0	0	1	0	1	0	0	0	1	0	1	1	0	0	0	19									
	<i>FimK R</i>	0	0	1	0	1	0	1	0	1	0	2	1	0	0	0	30									
<i>P. a.</i>	<i>ETA F</i>	0	1	1	1	1	0	0	0	0	0	2	1	0	0	0	37									
	<i>ETA R</i>	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	11									
<i>P. m.</i>	<i>ZapA F</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0									
	<i>ZapA R</i>	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3									
	<i>UreCF.</i>	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	6									
	<i>UreC R</i>	0	0	1	0	1	0	0	0	1	0	2	1	0	0	0	29									
<i>E. c.</i>	<i>FimHF</i>	0	0	0	0	0	0	0	1	1	0	2	1	0	0	0	26	0	7	1	7	1	1	397	Primer-BLAST*	
	<i>FimHR</i>	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	6									
<i>K. p.</i>	<i>FimKF</i>	0	0	1	0	1	0	0	1	1	0	0	0	0	0	1	13									
	<i>FimKR</i>	0	0	0	0	0	0	0	1	1	0	1	1	0	0	1	17									
<i>P. a.</i>	<i>ETAF</i>	0	0	1	1	1	0	0	1	1	0	1	1	0	0	1	33									
	<i>ETAR</i>	0	0	1	1	1	0	0	0	1	0	1	1	0	0	1	30									
<i>P. m.</i>	<i>ZapAF</i>	0	0	1	0	1	0	0	1	1	0	1	1	0	0	0	22									
	<i>ZapAR</i>	0	0	1	0	1	0	0	1	1	0	2	1	0	0	0	32									
	<i>UreCF</i>	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	12									
	<i>UreCR</i>	0	0	1	0	1	0	0	0	0	0	1	1	0	0	0	16									
<i>E. c.</i>	<i>FimH F</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	29	This study	
	<i>FimH R</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0									
<i>K. p.</i>	<i>FimK F</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1									
	<i>FimK R</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0									
<i>P. a.</i>	<i>ETA F</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0									
	<i>ETA R</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0									
<i>P. m.</i>	<i>ZapA F</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0									
	<i>ZapA R</i>	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3									
	<i>UreCF.</i>	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3									
	<i>UreC R</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	2								

F: forward; R: reverse. Ec: *E. coli*. Kp: *K. pneumoniae*. Pa: *P. aeruginosa*. Pm: *P. mirabilis*. Ss: *S. saprophyticus*. Ef: *E. faecalis*. * <http://www.ncbi.nlm.nih.gov/tools/primer-blast/>.

general application of this computational tool by non-AG experts. It is worth highlighting that the AG results are dependent on these values. In this work, we implemented an adjusted hybrid algorithm of Particle Swarm Optimization-Simplex [7], the Mult-PSOS, in order to obtain an autoconfigured computational strategy, which allows the rapid and reliable prediction of a set of oligonucleotides for simultaneous amplification of multiple genes in a single PCR (Multiplex-PCR). This strategy consists of the selection of a first primer pair (and subsequent pairs), following all the restrictions required for an adequate primer design. In the optimization process, the parameters and restrictions were defined and weighted in accordance with their

relevance in the primer performance (specificity and sensitivity) in a Multiplex-PCR. In this way, the most penalized restriction violations were the appearance of the primer sequence more than once in a target DNA (Uni(Pt)), and the size difference among amplicons (ladM(Pt)). The second highest penalizing value was given to the factors that affect the sensitivity of the oligonucleotides, such as hairpin loop structures (self-complementary sequences) and Ta and Tm differences greater than 5°C among primers.

To evaluate the accuracy of the method, an *in vitro* test with the Mult-PSOS predicted primers was made taking as targets five gene sequences, previously described in a Multiplex-PCR for simultaneous identification of the

bacteria most prevalently implicated in UTI [20]. Two additional genes were added for the Multiplex-PCR design from *S. saprophyticus* and *E. faecalis*, due to their prevalence in UTI etiology [10, 12, 17]. Using the restriction parameters, 14 oligonucleotide sequences were selected by the algorithm as the optimal sense and anti-sense primers. In the *in vitro* Multiplex-PCR, amplification of all the target sequences was achieved in only two tests, validating the good performance of the Mult-PSOS. It is worth mentioning that in the first assay, five of the six amplicons (*K. pneumoniae*, *P. aeruginosa*, *P. mirabilis*, *S. saprophyticus*, and *E. faecalis*) were successfully replicated, and the absence of one of the target genes (*E. coli*) was attributed to reduced molecular weight and to poor quality of the DNA template. A second test of Multiplex-PCR was performed with the same mixture and cycle conditions as the previous assay, but with *E. coli* genetic material obtained by the boiling-DNA extraction method. With this modification, 100% of the amplicons were amplified.

In general, the replication efficiency was proportional to the molecular weight of the amplicon, with the lowest rate being associated with the *FimH* gene from *E. coli* (235 bp). These kinds of variations are very common in Multiplex-PCR assays. Moreover, in this work, they were principally associated with the quality of the DNA template and the primer availability in the reaction [24]. The specificity of the primer sequences was verified *in silico* by nucleotide BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_SPEC=WGS&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch) against human, bacteria, fungal, and protozoan databases. The results showed non-significant cross homology regions in the primers. The specificity of the Mult-PSOS predicted primers was also confirmed *in vitro*, using DNA samples from *S. marcescens*, *Salmonella* spp., and *S. aureus* without any band amplification.

The primer sequences designed by the Mult-PSOS satisfied all the mandatory restrictions (see Eqs. (1)–(4)) for the adequate performance of Singleplex and Multiplex-PCRs (Eqs. (5)–(18)). The adequate adjustment of the parameters was confirmed by the Gene Runner 3.01 free software (<http://www.generunner.net/>) and the online tool Multalin (<http://multalin.toulouse.inra.fr/multalin/multalin.html>). In accordance with the results, we can conclude that Mult-PSOS is an autoconfigured algorithm that works in a sequential manner, facilitating the efficient prediction and selection of the primer sequences for the simultaneous replication of up to seven genes. The optimization of the reaction conditions for the adequate amplification of all amplicons was accomplished in a single step (with the

correct amount of DNA template and primers), decreasing significantly the need for the multiple testing by trial and error. This tool can be used for the systematic design and implementation of Multiplex-PCRs for clinical and forensic applications and can be employed by non-expert users in computational techniques or primer design problems.

Acknowledgments

This study was supported by grants from the Vicerrectoría de Investigaciones-Universidad de Santander (Project Number 034-11). We would like to thank Emile Blanchette for assistance in the proofreading of this paper.

References

1. Abd-Elsalam KA. 2003. Bioinformatic tools and guideline for PCR primer design. *Afr. J. Biotechnol.* **2**: 91-95.
2. Abdeldaim GM, Strålin K, Korsgaard J, Blomberg J, Welinder-Olsson C, Herrmann B. 2010. Multiplex quantitative PCR for detection of lower respiratory tract infection and meningitis caused by *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Neisseria meningitidis*. *BMC Microbiol.* **10**: 310.
3. Allawi HT, SantaLucia J. 1997. Thermodynamics and NMR of internal GT mismatches in DNA. *Biochemistry* **36**: 10581-10594.
4. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403-410.
5. Amghalia E, Nagi AA, Shamsudin MN, Radu S, Rosli R, Neela V, Rahim RA. 2009. Multiplex PCR assays for the detection of clinically relevant antibiotic resistance genes in *Staphylococcus aureus* isolated from Malaysian hospitals. *Res. J. Biol. Sci.* **4**: 444-448.
6. Apte A, Daniel S. 2009. PCR primer design. *Cold Spring Harb. Protoc.* **2009**: pdb-ip65.
7. Begambre O, Laier JE. 2009. A hybrid particle swarm optimization – simplex algorithm (PSOS) for structural damage identification. *Adv. Eng. Softw.* **40**: 883-891.
8. Boehme CC, Nabeta P, Hillemann D, Nicol MP, Shenai S, Krapp F, et al. 2010. Rapid molecular detection of tuberculosis and rifampin resistance. *N. Engl. J. Med.* **363**: 1005-1015.
9. Cheng YH, Chuang LY, Yang CH. 2012. Fuzzy adaptive particle swarm optimization for confronting two-pair primer design. In: *Proceedings of the International MultiConference of Engineers and Computer Scientists. IMECS 2012* **1**.
10. Foxman B. 2002. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Am. J. Med.* **113**: 5-13.
11. Gopinath K, Singh S. 2009. Multiplex PCR assay for simultaneous detection and differentiation of *Mycobacterium tuberculosis*, *Mycobacterium avium* complexes and other *Mycobacterium* species directly from clinical specimens. *J.*

- Appl. Microbiol.* **107**: 425-435.
12. Grabe M, Bishop MC, Bjerklund-Johansen TE, Botto H, Cek M, Lobel B, et al. 2010. *Guidelines on Urological Infections. European Association of Urology.* **110**.
 13. Gupta S, Bandyopadhyay D, Paine SK, Gupta S, Banerjee S, Bhattacharya S, Bhattacharya B. 2010. Rapid identification of *Mycobacterium* species with the aid of multiplex polymerase chain reaction (PCR) from clinical isolates. *Open Microbiol. J.* **4**: 93.
 14. Johnson JR, Stell AL. 2000. Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. *J. Infect. Dis.* **181**: 261-272.
 15. Kennedy J, Eberhart R. 1995. Particle swarm optimization. *Proc. IEEE Int. Conf. Neural Networks* **4**: 1942-1948.
 16. Liarte DB, Murta SM, Steindel M, Romanha AJ. 2009. *Trypanosoma cruzi*: multiplex PCR to detect and classify strains according to groups I and II. *Exp. Parasitol.* **123**: 283-291.
 17. Martineau F, Picard FJ, Ménard C, Roy PH, Ouellette M, Bergeron MG. 2000. Development of a rapid PCR assay specific for *Staphylococcus saprophyticus* and application to direct detection from urine samples. *J. Clin. Microbiol.* **38**: 3280-3284.
 18. Mullis KB, Faloona FA. 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155**: 335.
 19. Nelder JA, Mead R. 1965. A simplex method for function minimization. *Comp. J.* **7**: 308-313.
 20. Padmavathy B, Vinoth KR, Ameer P, Deepika SS, Vaidehi T, Jaffar ABM. 2012. Rapid and sensitive detection of major uropathogens in a single-pot multiplex PCR assay. *Curr. Microbiol.* **65**: 44-53.
 21. Rachlin J, Ding C, Cantor C, Kasif S. 2005. MuPlex: multi-objective multiplex PCR assay design. *Nucl. Acids Res.* **33**: 544-547.
 22. Reddington K, O'Grady J, Dorai-Raj S, Maher M, Van Soolingen D, Barry T. 2011. Novel multiplex real-time PCR diagnostic assay for identification and differentiation of *Mycobacterium tuberculosis*, *Mycobacterium canettii*, and *Mycobacterium tuberculosis* complex strains. *J. Clin. Microbiol.* **49**: 651-657.
 23. Shen Q, Shi W-M, Kong W. 2008. Hybrid particle swarm optimization and tabu search approach for selecting gene for tumor classification using gene expression data. *Comput. Biol. Chem.* **32**: 53-60.
 24. Sibley CD, Peirano G, Church DL. 2012. Molecular methods for pathogen and microbial community detection and characterization: current and potential application in diagnostic microbiology. *Infect. Genet. Evol.* **12**: 505-521.
 25. Stankowska D, Kwinkowski M, Wieslaw K. 2008. Quantification of *Proteus mirabilis* virulence factors and modulation by acylated homoserine lactones. *J. Microbiol. Immunol. Infect. Dis.* **41**: 243-253.
 26. Tsalik EL, Jones D, Nicholson B, Waring L, Liesenfeld O, Park LP, Woods CW. 2010. Multiplex PCR to diagnose bloodstream infections in patients admitted from the emergency department with sepsis. *J. Clin. Microbiol.* **48**: 26-33.
 27. Weissensteiner T, Nolan T, Bustin SA, Griffin HG, Griffin A. 2010. *PCR Technology: Current Innovations*, 2nd Ed. CRC Press, Boca Raton, FL.
 28. Wu JS, Lee C, Wu CC, Shiue YL. 2004. Primer design using genetic algorithm. *Bioinformatics* **20**: 1710-1717.
 29. Wu LC, Horng JT, Huang HY, Lin FM, Huang HD, Tsai MF. 2007. Primer design for multiplex PCR using a genetic algorithm. *Soft Comput.* **11**: 855-863.
 30. Xiao L. 2010. Molecular epidemiology of cryptosporidiosis: an update. *Exp. Parasitol.* **124**: 80-89.
 31. Yang CH, Cheng YH, Chuang LY, Chang HW. 2009. Genetic algorithm for the design of confronting two-pair primers, pp. 242-247 In: *Bioinformatics and BioEngineering, 2009. BIBE'09. Ninth IEEE International Conference.*
 32. Yang CH, Lin MC, Chuang LY. 2010. Primer design for the PCR in methylation studies using PSO. *Proceedings of the International MultiConference of Engineers and Computer Scientists.* **1**.
 33. Yang CH, Chang HW, Ho CH, Chou YC, Chuang LY. 2011. Conserved PCR primer set designing for closely-related species to complete mitochondrial genome sequencing using a sliding window-based PSO algorithm. *PLoS One* **6**: 17729.