

Research Article

In Silico Mining of Simple Sequence Repeats in Whole Genome of *Xanthomonas sp.*

Mukti Jaiswal and Anjana Pandey*

Nanotechnology and Molecular Biology laboratory, Department of Biotechnology, University of Allahabad, Allahabad-211002, Uttar Pradesh, India

Abstract

In current scenario, microsatellites are a large source of genetic markers. In this study, we mined simple sequence repeats in whole genome of *Xanthomonas* species (*Xanthomonas axonopodis* pv. citri, *Xanthomonas oryzae* pv. *oryzae*, *Xanthomonas campestris* pv. *Campestris*) by in silico methods. A total of 640 SSRs, 377 SSRs and 541 SSRs were detected in whole genome having density of 1SSR/8.08 kb, 1SSR/13.10 kb and 1SSR/9.510 kb for 5175.554 kb, 4941.439 kb and 5148.708 kb sequences length respectively. The results elucidated, only 32 types (0.618%), 39 types (0.789%) and 96 types (1.864%) of SSR sequences were present in *Xanthomonas axonopodis* pv. *citri, Xanthomonas oryzae* pv.oryzae and Xanthomonas campestris pv. campestis respectively. Depending on the repeat units, the length of SSRs ranged from 10 to 18 bp for di-, 12 to 27 bp for tri-, 12 to 24 bp for tetra-, 20 to 35 bp for penta-, 24 to 72 for hexa-, 21 to 133 bp for hepta-, 24 to 40 bp for octa- and 27 to 36 bp for nano nucleotide repeats. Di-nucleotide repeats were the most frequent repeat type (70.97%) followed by tri-nucleotide (22.23%), hepta-nucleotide (2.86%), hexa-nucleotide (1.91%) and tetra-nucleotide (1.392%) in all three species of *Xanthomonas*. SSR containing sequences of *Xanthomonas* species could not assign any specific class of protein (77.56%) due to the absence of homologs in the protein sequence database and these could be treated as an ideal molecular marker.

Keywords: *Xanthomonas*; Microsatellite; Molecular marker; Genome annotation; Protein sequence database; Simple sequence repeat

Introduction

Simple sequence repeats (SSRs), or microsatellites or short tandem repeats are short repeat motifs (1-6 bp) that are present in both protein coding and non-coding regions of DNA sequences [1-3]. SSRs are highly abundant and exhibit extensive levels of polymorphisms in prokaryotic genomes [4] and show a high level of length polymorphism due to insertion or deletion mutations of one or more repeat type [5]. SSRs being more abundant in noncoding regions than in exons because the lacks of selective constraints prevent correction mutations at these alleles [6,7]. Moreover, different taxon varies in abundance of different types of SSRs [8]. Strand slippage replication is generally considered to be the primary mechanism for the generation of microsatellite polymorphisms [9-12]. Recombination may also act on these sequences by changing repeat number through unequal crossover or gene conversion. The SSRs are either mined conventionally [13-17] or database sequences of genome and Expressed sequence tags (ESTs), which represent the expressed part of genome also serve as source of SSRs [18-20].

SSRs fulfils the criteria of Ideal molecular markers which are highly polymorphic, provide reproducible results and be simple to assay [4]. It has been also found useful in numerous DNA- and PCR-fingerprinting experiments for strain typing of a variety of fungi without prior knowledge of their abundance and distribution in the investigated fungal genomes [21,22] and are also useful across a number of related plant species [23,24]. Because of their high mutability, SSRs are thought to play an active role in genome evolution by creating and maintaining genetic variation [5]. The genus *Xanthomonas* is a diverse and economically important group of bacterial phytopathogens, belonging to the gamma-subdivision of the proteobacteria. *Xanthomonas axonopodis pv. citri (syn. X. citri pv. Citri*) [25], which is a bacteria, causes asiatic citrus canker and reduces fruit quality and yield. *C. paradisi* (grapefruit) and *C. aurantifolia* (Mexican lime) are the most

susceptible in field. C. reticulata (mandarin/tangerine) and C. sinensis (sweet orange) are relatively tolerant [26,27]. Significantly, no citrus species is resistant to XAC after artificial inoculation, suggesting that there is no true genetic resistance against XAC and that field tolerance is mainly due to variation in growth habit [28]. Xanthomonas oryzae *pv. oryzae* causes bacterial blight by invading the vascular tissue, which constrain production of this staple crop in much of Asia and parts of Africa. Tremendous progress has been made in characterizing the diseases and breeding for resistance. Xanthomonas oryzae pv. oryzae is important, from the standpoint of food security and as models for understanding fundamental aspects of bacterial interactions with plants. Xanthomonas campestris pv. campestris (Xcc), a gram negative aerobic rod, is the causal agent of black rot, which affects crucifers such as Brassica and Arabidopsis. The Xcc bacterium also infects weeds, including Arabidopsis thaliana, which has been sequenced and is the model species used in plant research.

The present study has been conducted for mining of whole genome of Xanthomonas species (Xanthomonas axonopodis pv. citri, Xanthomonas oryzae pv. oryzae, Xanthomonas campestris pv. campestris) available at National centre for biotechnology information (NCBI) to find out the distribution and abundance of SSRs for the development of markers and to annotate SSR containing sequences. In the present study, we performed comparison of the complete genome of three different species of Xanthomonas (Xanthomonas axonopodis

*Corresponding author: Anjana Pandey, Department of Biotechnology, University of Allahabad, Allahabad -211002, Uttar Pradesh, India, Tel: +91-9453029719; E-mail: mukti.jaiswal@gmail.com

Received July 14, 2014; Accepted August 18, 2014; Published August 20, 2014

Citation: Jaiswal M, Pandey A (2014) *In Silico* Mining of Simple Sequence Repeats in Whole Genome of *Xanthomonas sp.* J Comput Sci Syst Biol 7: 203-208. doi:10.4172/jcsb.1000157

Copyright: © 2014 Jaiswal M, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

pv. citri str. 306, Xanthomonas oryzae *pv. oryzae* KACC10331 and Xanthomonas campestris *pv. campestris str.* 8004 chromosome with each other. This comparative genomics approach has greatly accelerated the study of the molecular basis of pathogenicity and virulence of Xac.

Materials and Methods

Retrieval of whole genome sequences

Complete genome sequences for the above three species of *Xanthomonas axonopodis pv. citri, Xanthomonas oryzae pv. oryzae and Xanthomonas campestris pv. campestris* were downloaded from National centre for Biotechnology Information (NCBI) Gen Bank having accession numbers AE008923.1, AE013598.1 and CP000050.1 respectively.

Detection of SSRs

The harvesting of the SSRs was done using SSRIT (Simple sequence repeat identification tool) software [29]. The minimum length of SSR was fixed at 10 bp. The SSRs were defined as _10 bp di-nucleotide repeats; _12 bp trinucleotide repeats; _16 bp tetra-nucleotide repeats; _20 bp penta-nucleotide repeats; _24 bp hexa-nucleotide repeats; _21 bp hepta-nucleotide repeats; _24 bp octa-nucleotide repeats and; _27 bp nano-nucleotide repeats. The poly A and poly T repeats were removed by using Microsoft Word program, as these are not considered as SSRs, due to their presence at 3' ends of mRNA/cDNA sequences.

Annotation of SSR containing sequences

Functional annotation of all the SSR containing sequences was determined on the basis of 70% similarity against non-redundant (nr) protein database entries. It was performed using program Basic Local Alignment Search Tool (BLAST) and its variant BLASTX [30]. The resulting proteins obtained through similarity search by BLASTX program were classified into their respective classes.

Results

In the present study, whole genome of Xanthomonas axonopodis pv. citri, Xanthomonas oryzae pv. oryzae and Xanthomonas campestris pv. campestis sequences available at NCBI were searched for microsatellites with a minimum length of 10 bp. A total of 640 SSRs detected in whole genome of Xanthomonas axonopodis pv. citri, 377 SSRs in whole genome of Xanthomonas oryzae pv. oryzae and 541 SSRs detected in whole genome of Xanthomonas campestris pv. campestis has been screened, excluding Poly A and Poly T. Depending upon the length of the repeat unit itself (1-9 bp), the lengths of SSRs varied from 10 to 133 bp, respectively. Figure 1 shows the frequencies of SSRs with di-, tri-, tetra- penta-, hexa-, hepta-, octa-, and nanonucleotide repeat units. The most frequent repeat sequences type found within the whole genome of Xanthomonas axonopodis pv. citri, Xanthomonas oryzae pv. oryzae and Xanthomonas campestris pv. Campestis were dinucleotide repeats (70.97%) followed by tri -nucleotide (22.23%), hepta -nucleotide (2.86%), hexanucleotide (1.91%) and tetra-nucleotide (1.392%). Whereas, frequencies of repeats of penta-, octa- and nanowere 1.10%, 0.55% and 0.36% respectively in Xanthomonas campestris pv. Campestis. However no penta-, octa-and nano-nucleotide repeat was detected during the screening of Xanthomonas axonopodis pv. citri and Xanthomonas oryzae pv. oryzae. The average frequency percentages of all three Xanthomonas species are shown in Figure 1.

The observed frequency of different repeat types comprising the SSRs is presented in Figure 2A-2E and summarized in Table 1A and 1B SSRs were comprised of 5 types of di-nucleotide (GC)n, (CG)n, (TC)



n, (GT)n, (CT)n; 24 different types of tri-nucleotide (CCG)n, (CAG) n, (GGC)n, (CCA)n, (CGG)n, (ACG)n, (CGC)n, (GTG)n, (GCC) n, (GGT)n, (GTT)n, (ACC)n, (TGG)n, (GCG)n, (CTG)n, (CAA) n, (TGC)n, (AGC)n, (TGG)n, (TCC)n, (CAC)n, (GTG)n, (AGA)n and (GCA)n repeats; 19 types of tetra nucleotide (CGTG)n, (CGGC) n, (GAAG)n, (ATGC)n, (AGCG)n, (AGCC)n, (GCGG)n, (GCCC) n, (GCTG)n, (CCGG)n, (CCCG)n, (AAGC)n, (GGCA)n, (TGCC) n, (GGCT)n, (GCAC)n, (CGGG) n, (GCCA) n, (CACG) n repeats; 30 different types of hexa-nucleotide (CATCTA)n, (ACAGCG)n, (GTTGCG)n, (GTAGCG)n, (GGCAGT)n, (GGCAAT)n, (GCTGCC) (CAGGCC)n, n, (GGCGTT)n, (GTAGCT)n, (TTGGCT)n, (CAATGT)n, (GCATGG)n, (TGCTGT)n, (TTGCCG)n, (TTGGAA)n, (CATCTA)n, (ACACCA)n, (CCGCGG)n, (ATGGCC)n, (TCGGAA) (ATTGCC)n, (GTCATG)n, (GATGGA)n, (CGATAC)n, n, (ATGTCG)n, (CGCCAA)n, (TCGCTG)n, (TGTCGC)n, (AGCCAA)n repeats and 33 types of hepta-nucleotide (ATTGGCC)n, (CGGGAAT) n, (TGGGGAT)n, (TCGGGAA)n, (GGGATTC)n, (ACGCACA) AATCGGG)n, (GGGATTT)n, (GGGAGTC)n, (GGCGGAT) n, (TTCCCGA)n, (CGATTCC)n, (CGCAAAC)n, (CCAATCC) n, (CCGCTTG)n, (CAACCGC)n, (TAAGCAG)n, (ATCGGGA) n, (ATTCCCA)n, (TTCCCGC)n, (GGTTGCG)n, (CCGATTC) n, (GGGAATG)n, (GATTCGG)n, (GGATTCG)n, (GGGAATC) n. n, (GGATTGG)n, (GCGTGTC)n, (GGGAAGC)n, (GGACTGC)n, (GGGATTG)n, (GGGATGC)n, (GTTGCGT)n repeats.

Among di-nucleotide repeats, only (GC)n and (CG)n repeats were more frequent in all species of *Xanthomonas*. The most abundant repeats among trinucleotides were (CGC)n present in *X. campestris* followed by (GCC)n, (GCG)n, (CAC)n, (CCG)n, (CCA)n, (TGC) n, (GCC)n (CAA)n, (TGG)n, (GCA)n and others are present in less frequent repeats spread in all three genomes (Figure 2B). The most abundant repeat in tetra- nucleotide is (GAAG) present in *Xanthomonas* campestris and other repeat types were present in equal frequencies. Whereas, Hexa-nucleotide (ATGGCC) has two frequencies which is only found in *X. axonopodis pv. citri*. Other hexa-nucleotide repeats were present in equal frequencies. Hepta-nucleotides repeats had four frequency for SSRs (CGATTCC)n, three for (GGGAATG)n, and two for (GGGAATC)n which is only found in Xanthomonas campestris pv. campestris and rest of the SSRs were found in equal frequency.

Annotation of Xanthomonas axonopodis pv. citri, Xanthomonas oryzae pv. oryzae and Xanthomonas campestris pv. campestris sequences containing SSRs.

Citation: Jaiswal M, Pandey A (2014) *In Silico* Mining of Simple Sequence Repeats in Whole Genome of *Xanthomonas sp.* J Comput Sci Syst Biol 7: 203-208 doi:10.4172/jcsb.1000157



Parameters Values	X.citri	X.oryzae	X.campestris
Total number of SSRs identified	640	377	541
Number of sequences containing two SSRs	582	274	267
Number of sequences containing three SSRs	43	89	197
Number of sequences containing four SSRs	3	4	13
Number of sequences containing five SSRs	0	0	6
Number of sequences containing six SSRs	8	3	20
Number of sequences containing seven SSRs	4	7	33
Number of sequences containing eight SSRs	0	0	3
Number of sequences containing nine SSRs	0	0	2
Total number of sequences containing SSRs	32	39	96

 Table 1a:
 Summary of in silico mining of Xanthomonas axonopodis pv. citri,

 Xanthomonas oryzae pv. oryzae and Xanthomonas campestris pv. Campestris.

Repeat type	X.citri	X.oryzae	X.campestris
Dinucleotide	582 (90.90)*	274 (72.67)	267 (49.35)
Trinucleotide	43 (6.71)	89 (23.60)	197 (36.41)
Tetranucleotide	3 (0.46)	4 (1.06)	13 (2.40)
Pentanucleotide	0	0	6 (1.10)
Hexanucleotide	8 (1.25)	3 (0.79)	20 (3.69)
Heptanucleotide	4 (0.62)	7 (1.85)	33 (6.09)
Octanucleotide	0	0	3 (0.55)
Nanonucleotide	0	0	2 (0.36)
Total length of sequences searched (kb) circular DNA	5175.554	4941.439	5148.708
Density of SSRs	One/ 8.08	One/13.10	One/ 9.510

*Data in parentheses is the percentage value of the repeat type

 Table 1b:
 Summary of in silico mining of Xanthomonas axonopodis pv. citri,

 Xanthomonas oryzae pv. oryzae and Xanthomonas campestris pv. Campestris.

To determine the function of SSR containing sequences, the 32 sequences from Xanthomonas axonopodis pv. citri with SSRs were mined and annotated against the nr protein database available at http:// www.ncbi.nlm.nih.gov. For a small number, 4 (12.5%) SSR containing sequences, annotations were available (Figure 3A) of which 15 (5.49%) predicted proteins, 6 (2.19%) putative proteins, 124 (45.42%) hypothetical proteins and 128 (46.88%) belonged to different functional classes. Maximum sequences i.e. (28=87.5%) could not be assigned to any specific class due to the absence of a homologs in the protein sequence database. Likewise (39 sequences) SSRs of Xanthomonas oryzae pv. oryzae were mined and annotated against the nr protein database; for a small number i.e. 10 (25.64%) SSR containing sequence annotations available (Figure 3B) of which 59 (7.54%) predicted proteins, 22 (2.81%) putative proteins, 403 (51.53%) hypothetical proteins and 298 (38.10%) belonged to different functional classes. Maximum number of sequences i.e. (29=74.35%) could not be assigned to any specific class due to the absence of a homolog in the protein sequence database. For Xanthomonas campestris pv. campestris (96 sequences) SSRs were mined and annotated against the nr protein database; for small number of 28 (29.16) SSR containing sequence annotations available, (Figure 3C), in which 88 (8.76%) predicted proteins, 16 (1.59%) putative proteins, 498 (49.60%) hypothetical proteins and 402 (40.03%) belonged to different functional classes. Maximum number of sequences i.e. (68=70.83%) could not be assigned any specific class due to the absence of homologs in the protein sequence database and matched proteins were searched for SSRs but no protein found to contain SSR.

Discussion

In the present study, whole genome sequences of three species of

Xanthomonas retrieved from NCBI were mined for SSRs which could be used for designing the markers.

The abundance of the different repeats in the SSRs as detected in *Xanthomonas axonopodis* pv. *citri, Xanthomonas oryzae* pv. *oryzae* and *Xanthomonas campestris* pv. *campestris* were variable and not evenly distributed. These results are similar with earlier findings, which shows that the abundance of different repeats varied extensively depending upon the species examined [2]. We excluded poly A and poly T repeats due to which their number is under represented in the study. The SSRs with di-nucleotide repeats followed by tri-nucleotide were most abundant distribution in sequences of all three *Xanthomonas* species. Analysis of SSRs in terms of number showed hepta-nucleotides repeats present in largest quantity 33 (26.82) followed by hexa- 0 (24.39%), tri-23 (18.69%), tetra-19 (15.44%), Di-7 (5.6%), penta-6 (4.8%), octa-3 (2.43%) and nano-2(1.62%) respectively. In earlier studies on *E. coli* and Chlamydial strains characteristic SSR distribution with a marked relative abundance of tri- and hexanucleotide repeats are reported.



J Comput Sci Syst Biol ISSN: 0974-7230 JCSB, an open access journal Microsatellites contribute to the genomic variability of prokaryotes even in closely related genomes, as supported by the observed difference in SSR distribution patterns even between related genomes. It has been shown that Chlamydia-specific Pmp proteins are candidates of sequences where SSRs can contribute to genetic variation in the strains examined [31]. (AT)n and (CT)n are the most common repeat motif in fungi, plants and insects [32]. In our study (GC)n and (CG)n repeat were abundant while no (AT)n and (CT)n repeat were detected. The smaller repeat motifs were found to be predominant among SSRs identified and as the length of repeat unit increases their occurrence decreases. This may be because longer repeats have higher mutation rates, therefore less stable [33].

It has been proposed that numerous SSRs are the hot spots for recombination [34,35] especially di-nucleotide repeats are preferential sites for recombination due to their high affinity for recombination enzymes [36]. As molecular markers, dinucleotides are more important than the other SSRs and are one of the most sought-after markers because of their higher mutation rates. SSRs may affect DNA replication [4] and also plays important role in regulation of gene activity [37]. Some SSRs, found in upstream activation sequences, serve as binding sites for a variety of regulatory proteins [38,39]. In addition to this, the presence of repeated sequences within proteins has been detected in all organisms examined [40].

The protein coding sequences investigated, which are present in SSR sequence are as following: *Xanthomonas axonopodis* pv. *citri*, 4 SSR sequences i.e.12.5%, *Xanthomonas oryzae* pv. *oryzae* 10 SSR i.e.25.64% *Xanthomonas campestris* pv. *Campsites* 28 SSR i.e. 29.16%. The annotations for these protein coded by SSR sequence categorized into different classes of proteins (predicted, hypothical, putative and others). Due to the absence of a homolog in the protein sequence database, remaining SSRs sequences, i.e. 28 (87.5%), 29 (74.35%) and 68 (70.83%) corresponding to *Xanthomonas axonopodis* pv.*citri*, *Xanthomonas oryzae* pv. *oryzae* and *Xanthomonas campestris* pv. *campestris* respectively could not be assigned any specific class as well as no obvious function has yet been assigned.

Conclusion

Subsequently it can be concluded that the most frequent repeat sequences type found within the whole genome of *Xanthomonas axonopodis* pv. *citri*, *Xanthomonas oryzae* pv. *oryzae* and *Xanthomonas campestris* pv. *Campestis* were di-nucleotide repeats (70.97%) followed by tri-nucleotide (22.23%), hepta-nucleotide (2.86%), hexanucleotide (1.91%) and tetra-nucleotide (1.392%).

It can also be concluded that these non-coding SSR containing sequences of *Xanthomonas axonopodis* pv. *citri, Xanthomonas oryzae* pv. *oryzae* and *Xanthomonas campestris* pv. *Campestis* are 28 (87.5%), 29 (74.35%) and 68 (70.83%) respectively, would serve to be an ideal molecular marker. The current investigation is a valuable approach for saving both costs and time, provides quantitative data in understanding distribution of SSRs in the whole genome and information for designing of the molecular markers to be used in various studies.

Acknowledgment

Prof. Anjana Pandey of Department of Biotechnology, University of Allahabad, is gratefully acknowledged for continuous support and encouragement.

References

 Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996) Microsatellites in plants, a new class of molecular markers. Curr Sci 45: 45-54.

- Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10: 967-981.
- Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol 18: 1161-1167.
- Field D, Wills C (1996) Long, polymorphic microsatellites in simple organisms. Proc Biol Sci 263: 209-215.
- Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12: 4127-4138.
- Hancock JM (1995) The contribution of slippage-like processes to genome evolution. J Mol Evol 41: 1038-1047.
- Schlötterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20: 211-215.
- Hancock JM (1999) In: Microsatellite, Evolution and Application; Goldstein DB, Schlotterer C, Oxford University Press: Oxford, 1-9.
- Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4: 203-221.
- Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365: 274-276.
- 11. Zhu Y, Strassmann JE, Queller DC (2000) Insertions, substitutions, and the origin of microsatellites. Genet Res 76: 227-236.
- Nishikawa N, Nishikawa K (2002) A DNA sequence evolution analysis generalized by simulation and the Markov Chain Monte Carlo method implicates strand slippage in a majority of insertions and deletions. J Mol Evol 55: 706-711.
- Tautz D, Trick M, Dover GA (1986) Cryptic simplicity in DNA is a major source of genetic variation. Nature 322: 652-656.
- Panaud O, Chen X, McCouch SR (1995) Frequency of microsatellite sequences in rice (*Oryza sativa L.*). Genome 38: 1170-1176.
- Ahmad R, Struss D, Southwick SM (2003) Development and characterization of microsatellite markers in citrus. J Am Soc Horti Sci 128: 584-590.
- Kijas JMH, Fowler JCS, Thomas MR (1995) An evaluation of sequence tagged microsatellite site markers for genetic analysis within citrus and related species. Genome 38: 349-355.
- Kijas JMH, Thomas MR, Fowler JCS, Roose ML (1997) Integration of trinucleotide microsatellites into a linkage map of citrus. Theor Appl Genet 94: 701-706.
- Chen C, Zhou P, Choi YA, Huang S, Gmitter FG Jr (2006) Mining and characterizing microsatellites from citrus ESTs. Theor Appl Genet 112: 1248-1257.
- Liu Z, Tan G, Li P, Dunham RA (1999) Transcribed dinucleotide microsatellite and their associated genes from channel catfish (*Ictalurus punctatus*). Biochem Biophys Res Commun 259: 190-194.
- 20. Liu Z, Li P, Kocabas A, Karsi A, Ju Z (2001) Microsatellite-containing genes from the channel catfish brain: evidence of trinucleotide repeat expansion in the coding region of nucleotide excision repair gene RAD23B. Biochem Biophys Res Commun 289: 317-324.
- Meyer W, Castañeda A, Jackson S, Huynh M, Castañeda E; IberoAmerican Cryptococcal Study Group (2003) Molecular typing of IberoAmerican Cryptococcus neoformans isolates. Emerg Infect Dis 9: 189-195.
- Holton TA, Christopher JT, McClure L, Harker N, Henry RJ (2002) Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Mol Breed 9: 63-71.
- 24. Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet 106: 411-422.
- Gabriel DW, Kingsley MT, Hunter JE, Gottwald T (1989) Reinstatement of Xanthomonas citri (ex Hasse) and X. phaseoli (ex Smith) to species and reclassification of all X. campestris pv. citri strains. Intern J Syst Bacteriol 39: 14-22.

- Egel DS, Graham JH, Stall RE (1991) Genomic Relatedness of Xanthomonas campestris Strains Causing Diseases of Citrus. Appl Environ Microbiol 57: 2724-2730.
- 27. Gabriel DW, Hunter JE, Kingsley MT, Miller J, Lazo GR (1988) Clonal Population Structure of Xanthomonas campestris and Genetic Diversity Among Citrus Canker Strains. Mole Plant Micro Interac 1: 59-65.
- Brunings AM, Gabriel DW (2003) Xanthomonas citri: breaking the surface. Mol Plant Pathol 4: 141-157.
- 29. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, et al. (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa L*.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11: 1441-1452.
- Xiong J, Subramaniam S, Govindjee (1996) Modeling of the D1/D2 proteins and cofactors of the photosystem II reaction center: implications for herbicide and bicarbonate binding. Protein Sci 5: 2054-2073.
- Kassai-Jáger E, Ortutay C, Tóth G, Vellai T, Gáspári Z (2008) Distribution and evolution of short tandem repeats in closely related bacterial genomes. Gene 410: 18-25.
- Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21: 1111-1115.

- 33. Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146: 769-779.
- Jeffreys AJ, Murray J, Neumann R (1998) High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. Mol Cell 2: 267-273.
- Templeton AR, Clark AG, Weiss KM, Nickerson DA, Boerwinkle E, et al. (2000) Recombinational and mutational hotspots within the human lipoprotein lipase gene. Am J Hum Genet 66: 69-83.
- Biet E, Sun J, Dutreix M (1999) Conserved sequence preference in DNA binding among recombination proteins: an effect of ssDNA secondary structure. Nucleic Acids Res 27: 596-600.
- Sandaltzopoulos R, Mitchelmore C, Bonte E, Wall G, Becker PB (1995) Dual regulation of the Drosophila hsp26 promoter in vitro. Nucleic Acids Res 23: 2479-2487.
- Lue NF, Buchman AR, Kornberg RD (1989) Activation of yeast RNA polymerase II transcription by a thymidine-rich upstream element in vitro. Proc Natl Acad Sci U S A 86: 486-490.
- Csink AK, Henikoff S (1998) Something from nothing: the evolution and utility of satellite repeats. Trends Genet 14: 200-204.
- 40. Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D (1999) A census of protein repeats. J Mol Biol 293: 151-160.