

Research article

Open Access

Development of ESTs from chickpea roots and their use in diversity analysis of the *Cicer* genus

Hutokshi K Buhariwalla*^{†1}, Jayashree B^{†3}, K Eshwar³ and Jonathan H Crouch²

Address: ¹c/o JH. Crouch, Centro Internacional de Mejoramiento de Maiz Y trigo (CIMMYT), Apdo. postal 6-641, 06600 Mexico, D.F., Mexico, ²Centro Internacional de Mejoramiento de Maiz Y trigo (CIMMYT), Apdo. postal 6-641, 06600 Mexico, D.F., Mexico and ³International Crop Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh 502 324, India

Email: Hutokshi K Buhariwalla* - hutokshicrouch@yahoo.co.uk; Jayashree B - b.jayashree@cgiar.org; K Eshwar - k.eshwar@cgiar.org; Jonathan H Crouch - j.crouch@cgiar.org

* Corresponding author †Equal contributors

Published: 17 August 2005

Received: 29 November 2004

BMC Plant Biology 2005, 5:16 doi:10.1186/1471-2229-5-16

Accepted: 17 August 2005

This article is available from: <http://www.biomedcentral.com/1471-2229/5/16>

© 2005 Buhariwalla et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Chickpea is a major crop in many drier regions of the world where it is an important protein-rich food and an increasingly valuable traded commodity. The wild annual *Cicer* species are known to possess unique sources of resistance to pests and diseases, and tolerance to environmental stresses. However, there has been limited utilization of these wild species by chickpea breeding programs due to interspecific crossing barriers and deleterious linkage drag. Molecular genetic diversity analysis may help predict which accessions are most likely to produce fertile progeny when crossed with chickpea cultivars. While, trait-markers may provide an effective tool for breaking linkage drag. Although SSR markers are the assay of choice for marker-assisted selection of specific traits in conventional breeding populations, they may not provide reliable estimates of interspecific diversity, and may lose selective power in backcross programs based on interspecific introgressions. Thus, we have pursued the development of gene-based markers to resolve these problems and to provide candidate gene markers for QTL mapping of important agronomic traits.

Results: An EST library was constructed after subtractive suppressive hybridization (SSH) of root tissue from two very closely related chickpea genotypes (*Cicer arietinum*). A total of 106 EST-based markers were designed from 477 sequences with functional annotations and these were tested on *C. arietinum*. Forty-four EST markers were polymorphic when screened across nine *Cicer* species (including the cultigen). Parsimony and PCoA analysis of the resultant EST-marker dataset indicated that most accessions cluster in accordance with the previously defined classification of primary (*C. arietinum*, *C. echinospermum* and *C. reticulatum*), secondary (*C. pinnatifidum*, *C. bijugum* and *C. judaicum*), and tertiary (*C. yamashitae*, *C. chrossanicum* and *C. cuneatum*) gene-pools. A large proportion of EST alleles (45%) were only present in one or two of the accessions tested whilst the others were represented in up to twelve of the accessions tested.

Conclusion: Gene-based markers have proven to be effective tools for diversity analysis in *Cicer* and EST diversity analysis may be useful in identifying promising candidates for interspecific hybridization programs. The EST markers generated in this study have detected high levels of polymorphism amongst both common and rare alleles. This suggests that they would be useful for allele-mining of germplasm collections for identification of candidate accessions in the search for new sources of resistance to pests / diseases, and tolerance to abiotic stresses.

Background

Chickpea (*Cicer arietinum* L.) is one of the world's more important but less studied leguminous food crop with over 10 M ha grown across the Americas, the Mediterranean basin, East Africa, the Middle East, Asia and Australia [1]. As a grain legume it plays a significant role in the nutrition of the rural and urban poor in the developing world, as it provides a protein-rich supplement to cereal-based diets particularly of vegetarians and subsistence farmers who cannot afford meat. Despite its economic importance, chickpea productivity has been low because of yield losses to foliar and soil-borne fungal diseases (ascochyta blight, fusarium wilt and botrytis grey mould), insect pests (helioverpa pod borer) and abiotic stresses such as drought, cold and salinity. Sources of resistance and tolerance to these constraints exist in the wild *Cicer* germplasm yet remain largely unused by conventional breeding programs [2-5]. Three annual *Cicer* gene-pools have been defined based on the available hybridization reports, biochemical and molecular diversity analysis [6]: species within the primary gene-pool (*C. arietinum*, *C. reticulatum* and *C. echinospermum*) can be readily crossed usually generating fully fertile progeny; while species within the secondary gene-pool (*C. bijugum*, *C. pinnatifidum* and *C. judaicum*) can be successfully crossed with the cultigen *C. arietinum*, providing hybrid embryos are rescued. However, the progeny of crosses between primary and second gene-pools are frequently sterile; finally, species within the tertiary gene-pool (*C. cuneatum*, *C. chorasanicum*, *C. yamashitae* and others) have not yet been successfully crossed with the cultigen *Cicer arietinum*.

Cultivated chickpea has a relatively low level of diversity, due to a series of bottlenecks caused by restricted distribution of the wild progenitors and the founder effect associated with domestication [7]. In addition, chickpea breeding programs have limited themselves to a small number of cultivated genotypes having sources of biotic stress resistance and abiotic stress tolerance with little or no use of wild species [5]. This has resulted in limited durability of resistances to many of the major pests and diseases, and limited progress in abiotic stress tolerance breeding. Wild germplasm accessions held at the International Centre for Agricultural Research in Dry Areas (ICARDA) in Syria, the International Institute for the Semi-Arid Tropics (ICRISAT) in India and other genebanks in USA, Europe and Australia contain valuable sources of novel genetic variation for improvement of these traits [6]. The wild *Cicer* species from the secondary gene-pool (particularly *C. bijugum*, *C. pinnatifidum* and *C. judaicum*) are known to possess multiple sources of pest and disease resistance and tolerance to abiotic stresses including drought and cold [7,8].

Where species barriers can be overcome, molecular markers will facilitate rapid and efficient transfer of economically important traits into cultivated breeding pools. This will have the added advantage of broadening the narrow genetic base of this crop and thereby reducing its vulnerability to evolving pest and disease pressures, while simultaneously providing stable increases in yield far beyond the current 0.6% annual improvement [1].

Most types of molecular markers have been tested in chickpea including isozymes [9-12], restriction fragment length polymorphism (RFLP) [13,14] random amplified polymorphic DNA markers (RAPDs) [15,16] amplified fragment length polymorphisms (AFLPs) [17], sequence characterized amplified regions (SCARs) [18], inter-simple sequence repeat (ISSRs) [15], simple sequence repeat (STMS or SSR) [19-22], resistance gene analogs (RGAs) [23] DNA amplification fingerprinting (DAF) [24], and expressed sequence tags (ESTs) [25]. However, there is a low level of polymorphism detected in cultivated chickpea using isozyme/allozyme markers [9,10] and RFLP analysis [13,14]. In contrast, SSR markers have been shown to be highly polymorphic within the cultigen *C. arietinum* [26] and have been routinely utilized for creating genetic linkage maps [27-29]. These SSR markers have also been used as reference points for integration of the different chickpea linkage groups derived from inter- and intra-specific crosses [28,29]. To date there are over 500 SSR markers developed from chickpea genomic libraries [19,20,22]. In general, 30–50% of the chickpea SSR markers are polymorphic in any given breeding or intra-specific mapping population (HK Buhariwalla, unpublished data).

SSR markers remain the marker of choice for marker assisted selection in many breeding programs. However, SSR motifs may evolve too rapidly to be valuable as the sole assay for interspecific diversity analysis. In addition, markers shown to have tight genetic linkage to target genes in interspecific mapping populations may lose their selective power when used in backcross programs based on interspecific derivatives. Thus, there is a need for the development and utilization of gene-based markers that better serve molecular breeding applications and diversity analysis of germplasm.

In this context, the most important criteria for new markers are high reproducibility, detection of co-dominance polymorphism and suitability for rapid large-scale low cost screening. EST-based markers fulfill these criteria and since they are associated with the coding regions of the genome they also enhance molecular germplasm evaluation by capturing variation across transcribed regions and in genes of known function. This potentially resolves the problem of limited genomic coverage suffered by

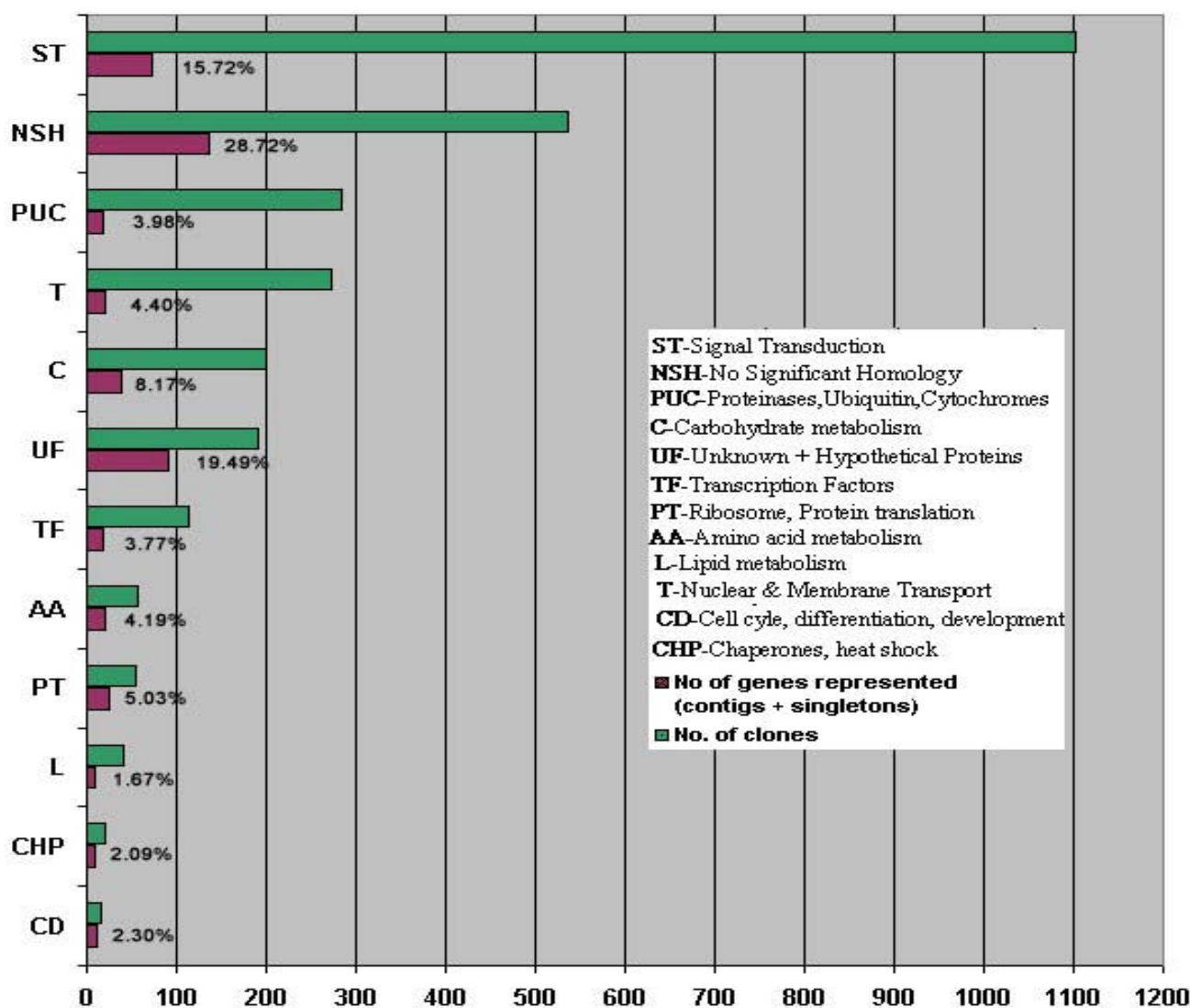


Figure 1

A summary of the number of EST clones and respective gene contigs classified in various functional categories based on alignments with public databases. Putative identifications were assigned to 2130 of the 2858 sequences generated from a chickpea root subtractive hybridization library. Percentages indicate the proportion of unigenes from the total number of unigenes identified in whole dataset.

traditional single gene phylogenetic studies [30,31]. Targeted EST development has already begun in chickpea, focusing on ABA-related mechanisms of water-deficit tolerance in epicotyl tissue [32].

Our objective in this study was to increase the public EST resource for chickpea with a particular focus on root tissue, as constitutive drought tolerance mechanisms (such as root system development) are a common target for the crop physiology community [33] and are increasingly

identified by genomics studies as important components of stress tolerance mechanisms [32,34]. Thus, an EST library was constructed after SSH of root tissue from two very closely related chickpea (*C. arietinum*) genotypes the landrace ICC 4958 and the popular local variety Annigeri, both considered to possess important sources of drought tolerance [35,36]. Following sequencing and functional annotation, we have generated over 100 EST markers which we have used in preliminary diversity analysis of wild and cultivated *Cicer* germplasm.

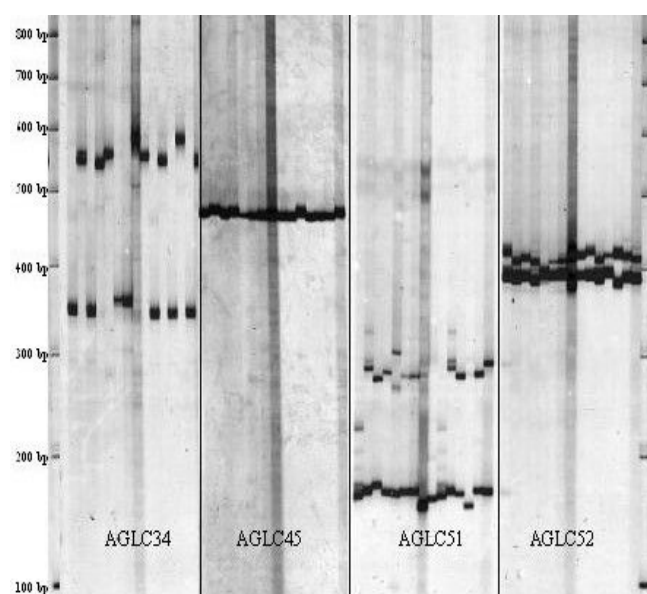


Figure 2
Polymorphic profiles of EST primers AGLC34, AGLC45, AGLC51 and AGLC52 screened on representatives of 8 wild species (IG69947, IG69960, IG69961, IG69974, IG69976, IG69986, IG69992, IG70029, ICC17116, ICC17121, ICC17122, ICC17126, ICC17141, ICC17148) and one cultivated genotype (ICC4958), separated by 6% non-denatured polyacrylamide gel electrophoresis and visualized by silver staining.

Results and discussion

The SSH process resulted in over 4000 clones expected to be largely associated with ICC 4958 (SSH tester) but could conceivably also include some highly constitutively expressed genes from Annigeri (SSH driver). Typically the cDNA inserts ranged from 0.5–1 kb. All clones were partially or fully sequenced through a single pass read from the 3' end and trimmed of vector and low quality sequences. A length threshold of 170 bp was set in view of the minimum expected size of a functional gene-encoding region (exon) [37]. Removal of short sequences (below 170 bp: largely comprising of problematic repeats and sequences predominantly composed of poly-A) resulted in 2858 high quality EST sequences with an average length of 480 bp all available at the ICRISAT Chickpea EST Database [38].

Analysis of ESTs and assembly of consensus sequences

Gene annotations were based on similarities to either known or putative ESTs in the public databases. All annotations were based on Blast searches, with a score threshold of ≥ 200 for BLASTn. For tBLASTx a score threshold of >100 was set, as these generally had e-values $<10^{-5}$ with a

minimum of 50% identity over at least 30% of the length of the protein, which are the commonly used thresholds for reliable sequence annotation [39,40]. Tentative functional annotations of EST sequences based on tBLASTx were grouped under 12 general categories based on the biochemical functions of the predicted proteins (Figure 1). The 2858 sequences were assembled into 210 contigs with 267 ESTs remaining unassociated (singletons). Putative identifications could be assigned to 2179 of the 2858 sequences (76%), representing 73% of the unigenes (178 singletons and 162 contigs), and all of these proteins were represented in other plant species. The 2858 sequences were assembled into 210 contigs with 267 ESTs remaining unassociated (singletons). Assuming that in most cases each contig represents one gene [41] a maximum number of 477 genes are represented in this dataset (available at NCBI: accession numbers CK148643–CK149150). The total length of contigs ranged from 240 to 1291 bases (with an average of 648 bases) while the largest number of ESTs assembled into one contig was 255. EST assembly leading to the generation of tentative consensus (TC) sequences has dramatically reduced the redundancy in this database, as reported elsewhere [42] (see Additional file 1).

Over three quarters of the 210 consensus sequences (77% equivalent to 164 contigs) were found to have significant similarity (tBLASTx >100) to sequences in public databases. Many of these annotations (47 consensus sequences) were validated through comparison with the CDD database using RPS-BLAST which compares a protein query sequence to a position-specific score matrix prepared from the underlying conserved protein domain alignment [43]. Conserved motifs consistent with tBLASTx annotations were identified in 47 consensus sequences providing an additional level of confidence for these assigned gene functions (see Additional file 2).

As expected, we observed a substantial number of the EST contigs and singletons to have no significant homology (NSH: 48 contigs and 89 singletons; 29%) to either nucleotide or protein sequences in public databases at the time of analysis. This compares well to the higher proportion of sequences (44%) reported with little or no homology from chickpea leaf tissue [44]. In addition, a substantial number of ESTs (approximately 19% of the unigenes) were found to have similarity with unannotated hypothetical, putative or unknown proteins contained in the public databases and were, therefore, placed in the unidentified function (UF) class.

PCR optimisation and unit costs

Optimization of PCR conditions is a critical precursor for accurate and robust large-scale marker screening and offers additional benefits for reducing the unit cost of gen-

Table 1: Passport data of *Cicer* accessions used in this study.

Accessions	Species	Alternate accession identifier	Biological status	Origin and location	Latitude	Longitude
ICC4958	<i>C. arietinum</i>	JGC 1	Cultivar	India	no data	no data
IG69947	<i>C. bijugum</i>		Wild	Turkey (Diyarbakir)	40.0	33.1
ICC17122	<i>C. bijugum</i>	ICCW 7	Wild	Turkey (Savur)	37.5	40.9
IG70029	<i>C. chorassanicum</i>		Wild	Afghanistan (Bamian)	67.5	34.5
ICC17141	<i>C. chorassanicum</i>	ICCW 26	Wild	Afghanistan (Shahidan)	34.2	69.7
IG69976	<i>C. cuneatum</i>		Wild	Ethiopia (Tigray)	38.5	14.1
IG69974	<i>C. echinospermum</i>		Wild	Turkey (Sanli Urfa)	30.4	37.4
IG69986	<i>C. judaicum</i>		Wild	Syria (Tartous)	35.5	35.0
ICC17148	<i>C. judaicum</i>	ICCW 33	Wild	Lebanon	no data	no data
IG69961	<i>C. pinnatifidum</i>		Wild	Turkey (Elazig)	39.2	38.4
ICC17126	<i>C. pinnatifidum</i>	ICCW 11	Wild	Turkey	38.7	39.3
IG69960	<i>C. reticulatum</i>		Wild	Turkey (Mardin)	40.6	37.3
ICC17121	<i>C. reticulatum</i>	ICCW 6	Wild	Turkey (Savur)	37.5	40.9
IG69992	<i>C. yamashitae</i>		Wild	Afghanistan (Kabul)	69.4	34.4
ICC17116	<i>C. yamashitae</i>	ICCW 1	Wild	Afghanistan (Shezghan)	34.7	69.7

otyping [45]. As optimization relies on the sequential investigation of each reaction variable, the process can take a long time, thus the conditions of optimization are rarely identified empirically and individual components are rarely all tested simultaneously. However, Cobb and Clarkson [46] devised a strategy in which the components of a PCR reaction could be tested together using the least number of experiments based on the Taguchi principle [47]. We adapted the Cobb and Clarkson [46], 9 reactions of 4 PCR components to a 5 reactions of 5 PCR components, the additional component being enzyme concentration. This further reduced the time and cost of optimization, whilst also minimizing the cost of the PCR screening by also considering enzyme concentration. Final primer (0.15 - 0.5 pmol) and Taq DNA polymerase (0.2 - 0.5 U) concentrations resulted in a reagent unit cost of US\$0.08 per PCR sample. This compares very favorable with other reports of PCR unit costs [48,49]. Clearly DNA extraction costs are an equally important component for molecular breeding programs and can also be greatly reduced as has been reported elsewhere [50]. The primer sequences and optimization parameters of all the polymorphic markers are given in Table 2. Optimization also resulted in fewer spurious amplification products and clearer polymorphic bands amongst the accessions studied (Figure 2).

EST marker diversity analysis

A total of 106 EST primers were designed, optimized and screened, of which 48 gave good amplification products in ICC 4958 (the SSH tester accession in the EST generation process) while 58 were deemed unsuitable as they either did not produce an amplification product or generated a complex pattern of bands which were difficult to evaluate. Of the 48 primers pairs screened across 1 cul-

tivated and 8 wild *Cicer* species, 44 were polymorphic (Table 2). A total of 167 polymorphic fragments ranging in size from 200 to >1000 bp was scored on polyacrylamide or agarose electrophoresis gels depending on product size (Figure 2). The number of alleles detected per primer ranged from 2 - 8 with the polymorphic information content (PIC) ranging 0.03 - 0.89 (Table 5).

Fourteen of the 106 EST's contained SSR motifs, of which 10 contained perfect trinucleotide repeats. A relatively large proportion of these EST-SSR markers (8) were derived from the no significant homology (NSH) and unknown protein functional (UF) classes, based on sequence comparisons (BLASTn tBLASTx and PSI-BLAST). Markers from these EST classes have readily detected polymorphism in the diverse germplasm tested (PIC 0.20-0.82). Since over 48% of the EST's generated from the chickpea root library fall into NSH and UF classes, these are clearly useful markers with potentially important, yet unknown functions.

Eleven of the polymorphic EST's were from stress annotated transcripts (AGLC2, AGLC16, AGLC 20, AGLC29, AGLC 34, AGLC45, AGLC52, AGLC53, AGLC55, AGLC66, AGLC68 and AGLC93); of these four contained SSR motifs (AGLC 34, AGLC55, AGLC67 and AGLC84). These stress related EST markers may be useful for allel mining of germplasm collections for the identification of candidate accessions with new sources of agronomically important traits and for candidate gene mapping [51].

The EST data-set generated in this study was used to compute pair-wise genetic distances between species according to Band similarity coefficient and UPGMA clustering. In order to display, with minimal distortion, the

Table 2: Characteristics of chickpea EST markers developed from EST root library with primer sequence, GenBank accession numbers, PCR optimized parameters including touch down amplification profile.

Marker	Primer Sequence (5'--3') Forward	Primer Sequence (5'--3') Reverse	GenBank accession numbers	Touch-down profile (°C)	PCR optimised reagents			
					Primer (pmol)	Mg ⁺⁺ (mM)	dNTP (mM)	Taq (u)
AGLC2	TGTCAGACTGAGCTGTGTATGAGA	TTGCCCGTATGGTTATGTTAGGAA	CK148953*	55-45	0.3	1.5	0.15	0.3
AGLC7	GACCCCCAAAAATGAAAAAGCA	TTGCCCATACATTCTTCAACCAA	CK148966	60-55	0.3	1.5	0.15	0.3
AGLC8	CAAACTCCTCAATAGCAGGCACA	GCTGTATCGGAGAGTGGTCAGA	CK149041*	55-45	0.3	1.5	0.15	0.3
AGLC9	ACTCCTGTAGTGGCATATCTTCGA	TGGTCCATTTATGCCGCTGGTA	CK148709	60-55	0.3	1.5	0.15	0.3
AGLC14	GCAGCAACTATTTACACTGGTA	CTCTCTGGGAGAAAGCTCGGAA	CK149086*	60-55	0.45	1.5	0.15	0.3
AGLC15	ACTGATCAAGGTCTCTTCTAGACA	CCCAACAACTGGACAAAGCAGA	CK149086*	60-55	0.45	1.5	0.15	0.3
AGLC16	GAGTACTTGCCAACTAGCTTAGGA	TTGGATATAACAGATGACGGGGAA	CK149087*	60-55	0.3	1.5	0.15	0.3
AGLC19	GCATCCTTCCCACTTCTTTGCA	GAATGGACTCGGATGTCTTAAGCA	CK148924*	60-55	0.3	1.5	0.15	0.3
AGLC20	AATGGTGATTCTGTCAGTCGCTA	CTGTCTGAAGAAAGTGAACGAA	CK148978*	60-55	0.3	1.5	0.15	0.3
AGLC27	CAAATTTCTGTTCTTCCACCCCAA	GGCGATCTTCGAGTCCATCGA	CK148933	60-55	0.3	1.5	0.15	0.3
AGLC28	GCTAAACCTTAGAGCAATGACTCA	CCTTGCTTGTGCCTTATCTTCCA	CK148934	60-55	0.3	1.5	0.15	0.3
AGLC29	TCTTCAACACCTCCATCTAACCTA	GACATGAAACCAAGCATCACA	CK148945*	60-55	0.3	1.5	0.15	0.3
AGLC34	CTTTACCAAAACACCTTCACCAA	TCTCTCTCTCTCTCTTCTGTTCCA	CK149140*	60-55	0.5	1.5	0.15	0.3
AGLC39	GCCGAGGTACACTTTACCAA	TCCTCACACTTCAGGTTCAACGA	CK149136	55-45	0.5	1.5	0.15	0.3
AGLC45	CTCCTCTTCTCCGTCTAGCA	CTGGTCTTGACGGGAGTGA	CK149070*	65-60	0.5	1.5	0.15	0.3
AGLC47	GTTTACATCATGACCCGCCCTA	TCACCAAGACCAGAACGTTCCA	CK148819	60-55	0.5	1.5	0.15	0.3
AGLC48	TGCCCAACGGTTTCTTTTACCA	TCAGAGATACTCGCCACCAA	CK148960	55-45	0.5	1.5	0.15	0.3
AGLC51	TCTTTGAGCAGCATTCATTCCACA	GAGTGCTACCTTCAAAGACTGCA	CK148677	55-45	0.5	1.5	0.15	0.3
AGLC52	CGATCAAGAACCAGTTTTGCAA	AAGATCGACAGGCGATCTGGTA	CK148718	55-45	0.5	1.5	0.15	0.3
AGLC53	CACCTCGCGTTCCGGTTCCA	CTGTCCATGCCCTTGTTCCA	CK148806	60-55	0.3	1.5	0.2	0.5
AGLC55	CAGGTCGCGTTGTTGCA	GGCCGAGGTACACTTTTCCA	CK149133*	60-55	0.5	2	0.1	0.3
AGLC57	TTCATCTGGCACTAGCATATCTGA	CGACAATCTTGCTTCAACAACCA	CK148900*	60-55	0.3	2	0.1	0.3
AGLC60	CATGTTTTCTACCCTCACAATGCA	TACTCACTTGTGTTCCAGACA	CK149124*	55-45	0.2	1	0.1	0.2
AGLC61	TTCGATCCTCCGACCCCGAA	TTGCTAGATCTGGATACTTCTCA	CK148802	60-55	0.3	2	0.1	0.3
AGLC64	TCTTCTTCTTCTTCTCAGCCACA	GTGGATTGGGAAATGTGAATGTCA	CK149128*	55-45	0.2	1	0.1	0.2
AGLC66	CCACAAAGGACGACAAACAACGA	CCCAACACGAACCACACGA	CK149070	60-55	0.2	2	0.2	0.5
AGLC67	ATCCATCACAACCTCAACTCA	CTCCGTCAACCTTCCGCAA	CK149102*	55-45	0.3	2	0.1	0.3
AGLC68	TGTTGTCTCGCCAATTCAAAGCA	CGTTTGGTGGCATTCTCTGCA	CK148906*	55-45	0.3	2	0.1	0.3
AGLC72	TTTAATTACGCGTTTCCACGA	GAAGACTTGAGACATGGGCACA	CK148928*	55-45	0.2	1	0.1	0.2
AGLC74	CGTGGGATTGAAAAAGTTGCTA	CACTACCAGCCAAAGCACTCA	CK149006*	55-45	0.2	1	0.1	0.2
AGLC75	CAACAACAACCTATCCGAACCTCA	ACTATCCCTAACCTTCCATCACC	CK148853	60-55	0.5	1	0.2	0.5
AGLC76	CATGAGTGGTAGTGGGAGTGGA	GTTTCGTTTGAGTCGTTTACTGGAA	CK148862	60-55	0.2	2	0.2	0.5
AGLC78	TCAACAACGCTACCCGATCCAA	TTCTCAAGAGCACCACAAAAGAGA	CK149132	60-55	0.3	2	0.1	0.3
AGLC79	CGGCGGCTATATTGGTTTGCA	TCCTAAACCCCACTTATCTCCCTA	CK149080*	60-55	0.3	2	0.1	0.3
AGLC82	TTTGATGGTCCGTCTCTCTCA	ACCGTCTCAGGATCAACTCGA	CK148871	60-55	0.3	2	0.1	0.3
AGLC83	TCTTCCGATCCTAAGAAAGAGCAA	ACCAATATGGAGAGCACCAGTCA	CK148894	60-55	0.3	2	0.1	0.3
AGLC84	CCACCTTCCATCTCCAATTCCAA	GACTGAATCGGAGAGGTTTCTCA	CK149089	60-55	0.15	2	0.1	0.2
AGLC85	CCAGCTTCTAATGTAGGTCTGCA	CAGCAGCAGCAGAGAGAGCA	CK148761	60-55	0.15	1	0.1	0.2
AGLC88	ACTTGGGCGTTCAAAATCTCA	CCATTACGATCAAGAGCTCAGGA	CK148768	55-45	0.3	2	0.1	0.3
AGLC93	TCCGAGCTGTGGATAGGGAA	GTTCCGCCTTCAATCCATGGAA	CK148904*	60-55	0.2	2	0.2	0.5
AGLC94	CCAATCTCCCTCATTCTTATTCCA	ACCAATTCCAAATTTCCAGCTCGA	CK149093	55-45	0.2	1	0.1	0.2
AGLC96	TCCATATGGCTGAGAACCCCAA	TTCTGAGGTTCAAGTAGTTCGGAA	CK149016*	60-55	0.2	2	0.2	0.5
AGLC98	CTCTTTCTTCCCTCTAGTTTCCA	CGGCGAACTCGTGTGTTGCTA	CK149016	55-45	0.2	1	0.1	0.2
AGLC101	TGTCCAAAATTGGGATCAGAGA	AGAACGACTTCAGCAGCAGCA	CK148993*	55-45	0.3	2	0.1	0.3

*Genbank accession numbers of the longest EST sequence of the contig, used in primer design.

genetic relationships between species, a Principle Coordinate Analysis (PCoA) was carried out. The first two dimensions of the PCoA plot indicate the presence of 5 clusters, which account for 20% and 17% of the total variation (Figure 3). Dollo and polymorphism parsimony analysis (DOLLOP) conducted on this data-set found one

most parsimonious tree with 301 steps (Figure 4) corresponding well with the PCoA analysis.

Both PCoA and parsimony analysis provide clear separation of most species. Both analysis grouped *C. echinospermum*, *C. reticulatum* and *C. arietinum*, together. These three

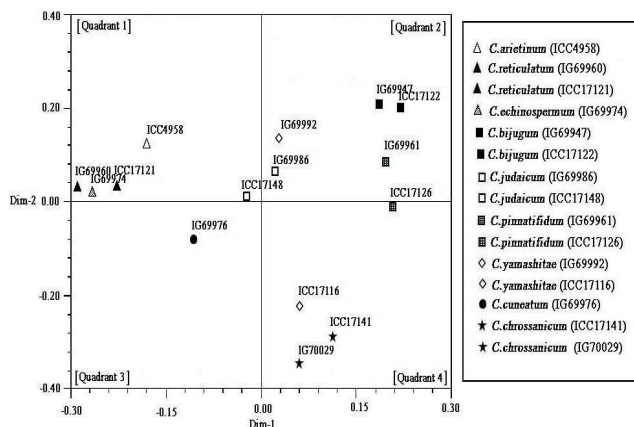


Figure 3
Principle Coordinate Analysis (PCoA) plot from diversity analysis of 44 EST markers screened across 14 accessions representing 8 wild *Cicer* species and one cultivated genotype (ICC4958).

species are classified together as the primary gene pool based on crossability studies. Similarly, all previous diversity studies based on allozymes [52], RAPD [16,53], microsatellite [54], ISSR [15,55] and AFLP [17] markers have also grouped *C. echinospermum*, *C. reticulatum* and *C. arietinum* together. The EST-based diversity analysis presented here shows *C. arietinum* (believed to be the wild annual progenitor [56] of cultivated chickpea) to be slightly distant to *C. echinospermum* and *C. reticulatum* but still within the same cluster (best reflected by the PCoA analysis Figure 3). All previous molecular diversity studies have clustered *C. bijugum*, *C. pinnatifidum* and *C. judaicum* together, although sometimes not well differentiated from the primary gene-pool species based on molecular diversity analysis. These three species are commonly considered to represent the secondary gene-pool as they require embryo rescue to recover viable progeny when crossed with the cultigen [6]. Both parsimony and PCoA analyses of the EST data-set clustered *C. judaicum*, *C. pinnatifidum* and *C. bijugum* together. This is in agreement with previous studies which demonstrated the close association of these three species. The current analysis suggests that *C. pinnatifidum* and *C. judaicum* are more closely related. This is in agreement with previous reports based on morphology analysis [56], seed storage protein analysis [57] and isozyme analysis [12]. In contrast, ecogeographical studies of the wild *Cicer* germplasm have suggested that *C. judaicum* to be quite diverged [7], which may explain the outlying position of one *C. judaicum* accession studied here (IG 69986) (Figure 4). Most other species generally fall into one of two clusters generally col-

lectively referred to as the tertiary gene pool (Figure 3 and Figure 4) from which no viable hybrids have been reported from crosses with the cultigen.

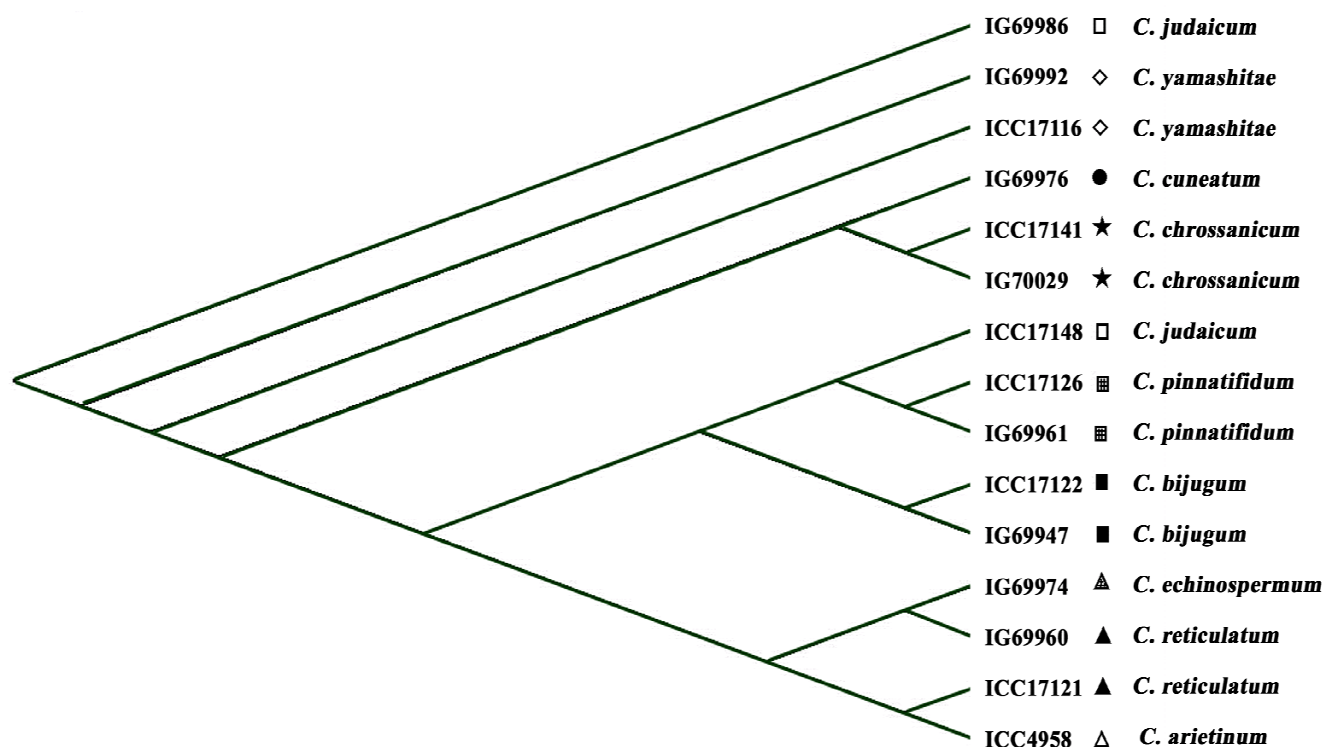
Parsimony analysis grouped *C. chorassanicum* with *C. cuneatum*, but placed the *C. yamashitae* accessions as outliers (Figure 4). This data supports the tertiary group proposed by Croser et al. [6] to consist of; *C. chorassanicum* which has never been successfully crossed with cultivated chickpea and *C. yamashitae* and *C. cuneatum* which have been crossed with the cultigen but the resultant progeny have failed to flower or proven sterile. However, this is not reflected in the PCoA (Figure 3).

The distribution of accessions from the same species within secondary and tertiary clusters suggests that these two gene-pools may not be as distinct as previously expected. This may indicate that there are some tertiary gene-pool species that might be more readily crossed with the cultigen than their taxonomic classification would otherwise indicate. However, since the current study is based on a limited numbers of accessions for both *C. cuneatum* and *C. yamashitae* further research is required to fully define the relationships between these species and the scope of new opportunities for plant breeders.

Conclusion

This is the first report of the use of EST-based markers for estimating genetic distances between annual *Cicer* species. We report on the development and characterization of 106 EST markers from chickpea sequences with functional annotations or unknown functions, some of which contain SSR motifs. These markers have detected high levels of polymorphism amongst the wild species studied here and initial results indicate that around 20% are polymorphic in intraspecific (*C. arietinum*) mapping populations, either directly or as cleaved amplified polymorphic sequences (CAPS) based markers [58].

The EST marker-based diversity analysis reported here broadly supports the *Cicer* taxonomy based on allozymic data and conclusions by Croser et al. [6]. The concurrence of our EST data with that of allozymic studies in particular suggests that many gene-based loci do have a common ancestry across *Cicer* species, as proposed but not substantiated by Choumane et al. [54]. Clustering patterns based on RAPD, AFLP and ISSR are generally very similar [15,17,55] but are substantially different to those based on the EST marker analysis reported here and the previously reported allozyme-based analyses [59]. This may be a consequence of RAPD, AFLP and ISSR markers sampling fundamentally different regions of the genome (under different evolutionary pressures) as compared with markers based on expressed genes. Unfortunately a detailed comparison of different assays is confounded by the vir-

**Figure 4**

Dendrogram of 14 *Cicer* accessions (representing 8 wild species and one cultivated genotype, ICC4958) based on parsimony analysis of 103 EST alleles.

tual absence of common accessions and the adoption of different statistical analyses in each study [6].

The EST database developed in this study provides a preliminary profile of some differentially expressed genes that may be associated with constitutive mechanisms important for stress tolerance and root development in chickpea roots. These include transcripts with putative annotations for proteases, T6P synthase, non-specific lipid transfer proteins, MRP-like ABC transporters, chaperones- HSP70, TCP-1-alpha, bZIP transcription factor, calcium ATPases, protein kinases, MRP4 glutathione-conjugate transporter, glutathione S-transferase, phosphoenol pyruvate carboxylase, S-adenosyl methionine synthetase (see Additional files 1 and 2 for full details). It is envisaged that these differentially expressed genes can be validated by the chickpea community and the most promising ESTs used in candidate gene mapping and allele mining. The resultant annotated EST markers will then be of substantial value for marker-assisted introgression programs based on interspecific crosses using wild *Cicer* species that harbour agronomically valuable genes.

One of the aims of chickpea breeding is to address the continuing need for cultivars adapted to particular geographical regions or with specific new ideotypes to ensure sustainability and profitability of production. An important means of continuing to achieve such breeding objectives is through the use of novel germplasm [60]. Many breeding programs, in particular those involving rice and wheat, have successfully utilized molecular and statistical approaches in accelerating the introduction of novel genes from wild species through marker accelerated back-cross breeding.

The availability of increasing amounts of sequence data in many legume species now offers the potential for routine development of gene-based markers. These provide the ultimate assay (or so-called 'perfect marker') for indirect trait selection and map-based cloning. Combining gene-based markers together with highly polymorphic flanking SSR markers will greatly assist in reducing linkage drag and increasing the speed and efficiency of subsequent introgression programs. Finally, molecular genetic diversity analysis based on EST markers may also provide a

Table 3: Permutations of PCR component concentrations used for optimization of PCR conditions for each marker.

Reactions	Primer (pmol)	Template (ng)	MgCl ₂ (mM)	dNTP (mM)	Enzyme (u)
1	0.2	5	1	0.1	0.2
2	0.2	15	2	1.5	0.5
3	0.3	5	1.5	0.2	0.5
4	0.3	10	2	0.1	0.3
5	0.5	10	1	0.2	0.3

means of predicting which accessions are most likely to produce fertile progeny when crossed with chickpea cultivars.

Methods

Plant material and growth conditions for SSH

Fifteen seeds from each chickpea accession (ICC 4958 and Annigeri) were sterilized with 25% chlorex (v/v) for 10 mins, rinsed twice with sterile distilled water, before placing in pre-sterilized plant pots containing sterilized vertisol. The soil was collected from the field, passed through a 2 mm sieve and autoclaved in plastic bags using a 60 min cycle at 120°C and repeated three times on consecutive days. Seeds were sown immediately after the last autoclave cycle.

Three seeds were sown per pot; the plants were maintained in a Conviron growth chamber (Conviron, Winnipeg, Man.) under optimal physiological conditions for chickpea growth: 25°C day (11 h) and 12°C night (13 h), 455 µmoles m⁻² sec⁻¹ illumination intensity, 80% day and 40% night relative humidity, and irrigation as required.

Roots were harvested 28 days after germination (with flowering expected after 30 days). The soil was washed off the roots in running water, the intact whole plant was lifted out of the soil and soil particles were gently rubbed off in running water with gloved hands. The root tissue was subsequently treated with (0.1% v/v) DEPC solution and immersed in sterile Milli Q water, excess water from the roots was blotted onto filter paper and roots were weighed (approx. 2 g), flash frozen in liquid nitrogen and stored at -80°C.

RNA isolation and suppressed subtractive hybridization

The subtraction of the two genomes was achieved through a process of PCR-based suppression of genes common to both genotypes, and genes that are differentially expressed are enriched. As the process is based on PCR, low copy number cDNA can be detected from the tester (ICC 4958). RNA isolation and library construction was carried out by Avestha Gengraine Technologies (Bangalore, India). Total RNA was isolated from the root tissue of ICC 4958 and

Annigeri using Trizol reagents (GIBCO-BRL) for RNA isolation, mRNA was purified using Oligotex (Qiagen GmbH, Hilden, Germany).

The suppression subtractive hybridization (SSH) process was carried out using a Clontech PCR-Select cDNA subtraction kit (Clontech INC. USA). Tester (ICC 4958) and driver (Annigeri) cDNA was restricted with *RsaI* for 90 mins and cDNA from the ICC 4958 was ligated with the required adaptors. Adaptor ligation was confirmed by PCR, followed by one round of over-night hybridisation with Annigeri as the driver and ICC 4958 as the tester. A second round of hybridisation was followed by primary PCR (27 cycles) and secondary PCR (12 cycles) according to the manufacturer's instructions for the PCR-Select. The PCR products of the subtraction were analyzed by gel electrophoresis and cloned using pCR4-TOPO cloning kit (Invitrogen Inc., USA). The cDNA clones were amplified in *E. coli* DH10B cells by electroporation and insert containing clones were selected by plating on LB agar containing 100 µg/ml ampicillin. Large colony forms were picked and used to regenerate single clone cultures in 96-well microtitre plates. After growth over-night at 37°C, glycerol was added to a final concentration of 15% and cultures were stored at -80°C.

EST sequencing

Cultures of transformed *E. coli* were grown over-night in 100 µl LB media containing 100 µg/ml ampicillin, the cells were centrifuged and suspended in sterile distilled water, heat denatured for 10 mins at 95°C, 1 µl of supernatant was used for insert amplification with M13 forward and reverse primers. Inserts (5 µl) were separated by electrophoresis on 1.2% agarose gel containing ethidium bromide to confirm amplification of a single insert and for quantification. After SAP/exonuclease I digest, 4 µl of the insert was sequenced using T₇ primer and 1/8 fraction of the recommended volume of dye terminator reagent (Big Dye terminator cycle sequencing kit v2, ABI-Foster City) for cycle sequencing. After removal of dye blobs by ethanol precipitation, DNA sequences were analyzed by capillary-electrophoresis using an Applied Biosystems genetic analyzer model ABI 3700 or ABI 3100.

Sequence analysis for predicting gene function

Base calling was performed using the ABI DNA Sequence Analysis Software (v3.7) or by Chromas v2.2 (Technelysium Pty Ltd. Australia). All sequences were scanned visually for quality of the peak shapes and corresponding base call. Vector sequences were trimmed using Sequencher v.4 (Gene Codes, Ann Arbor, MI) software and low quality bases (quality score <20) were trimmed from both ends of sequences. A total of 4000 sequences were analyzed, only sequences of more than 170 bp with less than 5% ambiguity were processed. All sequences containing interspersed or simple repeats were masked with the RepeatMasker software [61] using *Arabidopsis* as reference. The masked sequences were screened against the following DNA databases: *Arabidopsis thaliana*, *Medicago truncatula*, *Glycine max*, *Zea mays* and *Oryza sativa*, in the TIGR Gene Indices using default settings for BLASTn and tBLASTx searches [62]. The same sequences were also used to search the NCBI EST database (dbEST) using *Viridiplantae* as a limiter [63].

The comparison of chickpea sequences with those in public databases was carried out during 2002 – 2003, and blast searches with unique sequences were repeated in November 2004. An in-house script was used to retrieve best matches from the outputs of BLASTn and tBLASTx text files, retaining the following information: gene identifier, description, BLAST score, percent identity, alignment length, reading frame, and database name from the top-scoring alignments. Alignments with scores of ≥ 200 for BLASTn and ≥ 100 for tBLASTx were extracted into an MS-Access table. The individual ESTs were assembled into tentative consensus sequences (TCs) using Sequencher, the assembly parameters used for 'dirty data' were (minimum percentage match of 90 and minimum overlap 40 bp). A sequence was retained in a contig if it matched at least one sequence already in the contig based on pair-wise BLAST [63], with a percentage identity threshold of 90%. A consensus sequence was then derived for each cluster. Consensus sequences were translated into six possible reading frames with Transeq software [64], using the standard genetic code. The translated sequences were then submitted to Reverse Position Specific (RPS)-BLAST [63] to search for matches to the conserved domain database (CDD; all 4540-PSSMs) using default parameters. The RPS-BLAST outputs were similarly processed using an in-house script, retaining the following information from the top-scoring alignments: gene identifier, description, score and e-value. Each result was manually interpreted and classified according to the biochemical function of the predicted protein.

Plant material and DNA isolation

Fifteen chickpea accessions representing 9 annual *Cicer* species, including the cultivated species *C. arietinum*, were

used in this study (Table 1). Seeds of these accessions were germinated in Jiffy pots in a growth room and DNA was extracted from 4–5 pinnules of 15–20 day old individual plants with 2% CTAB extraction buffer containing 0.03% mercaptoethanol as described in [50]. DNA concentrations were determined by comparing the sample intensity (1 μ l) with that of known amounts of uncut lambda DNA by electrophoresis in 1.2% 'Ready-to-run' agarose gels (Amersham) containing ethidium bromide. DNA was diluted to a working stock concentration of 10 ng/ μ l and checked before use.

Primer design of EST markers

A total of 106 EST sequences were chosen that represented diverse functional/stress annotations, a selection of which contained SSR motifs. In addition, primers were also designed from a random selection of sequences categorized as having 'no significant homology' (NSH) and 'unknown protein function' (UF) when compared with either nucleotide or protein sequences in public databases.

Primer sequences were designed using Genefisher 1.1 software [65] using the following criteria 57 – 63°C melting temperature (T_m), 40–50% GC content, 17–24 bp primer length and the difference in T_m between forward and reverse primer pairs was limited to 2°C. For those EST sequences containing repeat motifs, primers were designed to flank the microsatellite region. Optimal primer pairs were selected using Netprimer, the primer design algorithm which analyzes all possible secondary structures including hairpins, self-dimers and cross-dimers [66]. The T_m and product length of 8 forward and reverse pairs per sequence from Genefisher software were evaluated by Netprimer software for minimal or no secondary structure. All primer and EST sequences are accessible through the ICRISAT Chickpea EST Database [38].

PCR optimization and evaluation

The chickpea accession ICC 4958 (from which the EST sequences were derived) was used as the reference accession for primer optimization. A total of 106 EST primers (synthesized by MWG, Germany) were optimized simultaneously for the following PCR components using a modified (5×5) grid [46]. On this basis initial optimization of the five major components in a PCR (concentrations of primer, template DNA, Mg^{++} , dNTP and enzyme) were empirically determined for each primer used in this study (Table 3). Optimal touch-down temperature and number of amplification cycle profiles were also determined for each primer pair (Table 4). This procedure is important to minimize non-specific amplification, maximize data accuracy and minimizes cost.

Table 4: PCR amplification profiles used for optimization of PCR conditions for each marker.

Name	Profile	Temperature°C	Time min. sec	Number of cycles
65-60	Denaturation	95	3.00	1
	Touch down	94	0.20	5
		65-60	0.20	
		72	0.30	
	Normal	94	0.20	30
		59	0.20	
		72	0.30	
		72	20.0	
	Extension	72	20.0	1
60-55	Denaturation	95	3.00	1
	Touch down	94	0.20	5
		60-55	0.20	
		72	0.30	
	Normal	94	0.20	30
		56	0.20	
		72	0.30	
		72	20.0	
	Extension	72	20.0	1
55-45	Denaturation	95	3.00	1
	Touch down	94	0.20	10
		55-45	0.20	
		72	0.30	
	Normal	94	0.20	30
		48	0.20	
		72	0.30	
		72	20.0	
	Extension	72	20.0	1

PCRs were performed in a total volume of 5 µl containing 5–10 ng of DNA and optimized concentration of the following: primer (0.15–0.3 pmoles), dNTP (0.1–0.2 mM), MgCl₂ (1.5–2 mM), Qiagen Taq (0.2–0.5 u) and 1 x buffer. PCR amplifications were carried out using a GeneAmp model 9700 thermocycler (Perkin Elmer-Applied Biosystems). As the annealing temperatures of the markers ranged from 47–65°C we opted to use 3 categories of "touchdown" temperature cycles (see Table 4).

PCR amplified products with a size range of 200–600 bp were resolved by non-denaturing polyacrylamide gels electrophoresis (6–10% acrylamide/bisacrylamide, 20:1 in TBE). Bands were visualized through a modified silver staining protocol [67] as follows: gels were immersed in water for 3 mins, followed by 20 mins in 0.1% CTAB solution and 0.3% ammonia solution for 15 mins. Silver staining solution was freshly prepared each day, consisting of 0.1 % (w/v) AgNO₃ in 4 mM NaOH solution, 0.5 to 0.6 ml of 25% ammonia was titrated until the cloudy suspension became clear. Gels were gently agitated in the silver nitrate solution for 30 mins, and developed in 1.5% (w/v) sodium carbonate and 0.02% (v/v) formamide

solution until bands appeared. The gels were rinsed in water, fixed in 1.5% glycerol solution and documented by scanning. PCR products above 600 bp were resolved by agarose gels electrophoresis (0.8–1%) containing ethidium bromide.

EST marker data collection and analysis

The amplification profile from each EST marker was checked for reproducibility and subsequently scored visually (and independently rescored) for the presence (1) or absence (0) of polymorphic bands. The degree of polymorphism was quantified using the polymorphic information content (PIC) calculator [68] based on the following formula:

$$PIC = 1 - \sum P_i^2 - \sum \sum P_{ij}^2$$

$$i = 1 \quad i = 1 \quad j = i+1$$

where P_i is the frequency of an individual genotype.

Pair-wise genetic distances were calculated with NTSYS-pc software version 2.0 [69] as $S_{xy} = 1 - (2n_{xy}/n_x + n_y)$ as $S_{xy} =$

Table 5: Characteristics of chickpea EST markers including SSR motif where present, annotation information (NSH indicates no significant homology), and, size range of amplification products, number of alleles and PIC (polymorphism information content) values based on screening of germplasm in Table 1.

Marker	SSR motif	Annotation	Product size (bp)	No. of alleles	PIC
AGLC2		Valyl tRNA synthetase	620	2	0.50
AGLC7		D-3-phosphoglycerate dehydrogenase	330–335	2	0.44
AGLC8		RNA helicase	525–550	5	0.78
AGLC9		Unknown protein	680	4	0.62
AGLC14		Hypothetical protein	520–600	2	0.45
AGLC15		Hypothetical protein	520–600	4	0.65
AGLC16		WD-repeat protein like	365–370	2	0.32
AGLC19		Formin-like protein	350–355	4	0.50
AGLC20		Type IIB calcium ATPase	230–235	3	0.61
AGLC27		Putative Scarecrow gene regulator	330–335	3	0.43
AGLC28		Type 2A protein phosphatase	335–350	3	0.86
AGLC29		Non-specific lipid transfer protein	325–400	6	0.64
AGLC34	AG(20)	Putative protein kinase	350–600	5	0.75
AGLC39	AG(17)	No significant homology	275–290	3	0.74
AGLC45		Probable cysteine proteinase (EC 3.4.22.-)	410–420	2	0.29
AGLC47		(NM_127184) putative beta-ketoacyl-CoA synthase	200–210	2	0.59
AGLC48		Putative peroxisome assembly factor-2 (NM_100181)	400–410	3	0.89
AGLC51		RNA binding protein	170–180	4	0.55
AGLC52		Protein kinase	390–410	6	0.89
AGLC53		Acyl- [acyl-carrier protein] desaturase	390–400	5	0.87
AGLC55	CT(17)	Putative protein kinase	225–275	4	0.89
AGLC57	GGA(3)	Unknown protein; protein	>1000	3	0.56
AGLC60	TTC(3)	Hypothetical protein	310–350	7	0.82
AGLC61	GAA(3)	Protein of unknown function	450–470	5	0.55
AGLC64	TTC(6)	No significant homology	270–360	4	0.20
AGLC66		Probable cysteine protease (EC 3.4.22.-)	500–650	8	0.63
AGLC67	TCT(4)	Tryptophan synthase beta-subunit (TSB2)	>1000	4	0.56
AGLC68		Lon protease homolog I precursor	220–230	3	0.39
AGLC72		Putative GrrI protein	230–250	4	0.62
AGLC74	TTTA(3)	Glycosyl hydrolase family 17	400–450	5	0.75
AGLC75	AAC(5)	No significant homology	200–230	3	0.48
AGLC76	AAG(3)	RGA2 protein <i>Arabidopsis thaliana</i>	440–480	3	0.71
AGLC78	CAA(5)	Translation initiation factor eIF3 – like protein	430–450	2	0.44
AGLC79		No significant homology	330–410	5	0.03
AGLC82		Unknown protein; protein id: AtIg80280.I	300–320	3	0.50
AGLC83		Putative valyl tRNA synthetase [<i>Oryza sativa</i> -japonica cultivar-group]	>1000	3	0.74
AGLC84	AAG(3)	DEAD/DEAH box RNA helicase	430–450	3	0.42
AGLC85		Emb trypsin inhibitor cme precursor	410–450	3	0.47
AGLC88		Chromomethylase CMT2 <i>Arabidopsis thaliana</i>	>1000	5	0.78
AGLC93		BZIP family transcription factor	>1000	2	0.51
AGLC94		T48 ankyrin-like protein	480–500	4	0.67
AGLC96		Hypothetical protein CAB95829.I	310–360	5	0.59
AGLC98	AG(19)	Hypothetical protein	405–450	5	0.78
AGLC101		No Significant homology	370–400	4	0.70

$1 - (2n_{xy}/n_x + n_y)$ based on band coefficient [70], where n_{xy} are shared bands amongst the individuals, n_x and n_y are the number of fragments exhibited by each individual. Distance based on Band coefficient is similar to the genetic distance (GD) formula by Nei and Li [71]: $GD = 1 - S_{xy}$. The 15×15 similarity matrix (based on Band coefficient) was subjected to sequential agglomerative hierar-

chical nested (SAHN) clustering using UPGMA (unweighted pair-group method analysis). Principle coordinate analysis based on these distance estimates was performed using NTSYS, using consecutive commands of 'DCENTRE' and 'EIGEN' in order to generate a PCoA plot which is more informative in highlighting differences between major taxonomic groups than dendrogram repre-

sentations. Maximum parsimony analysis was conducted using the PHYLIP program Dollo and polymorphism parsimony analysis version 3.63 [72] and the consensus tree viewed in tree view.

Authors' contributions

HKB coordinated library construction and sequencing of a large proportion of the ESTs, experimental design, marker analysis and manuscript development. JB coordinated and/or carried out all bioinformatic analyses and participated in all aspects of manuscript development. KE carried out all aspects of marker optimization and screening, and contributed to various aspects of experimental design and data analysis. JHC participated in all aspects of experimental design, data interpretation and manuscript development and provided overall coordination of the project. All authors read and approved all versions of the manuscript.

Additional material

Additional File 1

List of the most abundant EST from the chickpea root EST database. Number of chickpea ESTs forming tentative consensus sequences (TC) with corresponding TC number and where known, putative annotation*. Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2229-5-16-S1.doc]

Additional File 2

Table of chickpea unigene sequences classified through RPS-Blast. Functional categories, with corresponding genbank ID numbers, Blastx descriptors, domains identified through RPS-Blast and e-values of chickpea unigene sequences. Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2229-5-16-S2.doc]

Acknowledgements

This work was supported by ICRISAT core funds including earmarked unrestricted grants from the governments of UK, Japan and the European Union. The authors would like to thank Dr HC Sharma for providing access to chickpea material, Pratibha R. for technical assistance with initial EST marker screening, Dr HD Upadhyaya for provision of gene-bank passport data for accessions used in this study, Seetha K for lab management, PVNS Prasad and KDV Prasad for assistance with figures and to Drs R Serraj and R Folkertsma for helpful suggestions during manuscript development. We greatly appreciate suggestions from an anonymous reviewer concerning this manuscript.

References

1. FAOSTAT data 2004 [http://faostat.fao.org].
2. Erskine W, Siddique K, Khan T, Cowling W: **Utilization of grain legume diversity.** In *Plant genetic resources of legumes in the Mediterranean* Edited by: Maxted N, Bennett SJ. Kluwer Academic, The Netherlands; 2001:311-326.
3. Knight E, Brinsmead R, Fordyce M, Wood J, Kelly A, Harden S: **Use of wild relative *Cicer echinospermum* in chickpea improvement.** *Proceedings of the 12 th Australian Plant Breeding Conference. Perth Australia 2002*:150-154.
4. Muehlbauer FJ, Kaiser WJ, Simon CJ: **Potential for wild species in cool season food legume breeding.** *Euphytica* 1994, **73**:109-114.
5. Singh KB, Malhotra RS, Halila H, Knight EJ, Verma MM: **Current status and future strategy in breeding chickpea for resistance to abiotic and biotic stresses.** *Euphytica* 1994, **73**:137-149.
6. Croser JS, Ahmad F, Clarke HJ, Siddique KHM: **Utilisation of wild *Cicer* in chickpea improvement- progress, constraints and prospects.** *Aust J Agric Res* 2003, **54**:429-444.
7. Berger J, Abbo S, Turner NC: **Ecogeography of annual wild *Cicer* species: the poor state of the world collection.** *Crop Sci* 2003, **43**:1076-1090.
8. Singh KB, Ocampo B, Robertson LD: **Diversity for abiotic and biotic stress resistance in the wild annual *Cicer* species.** *Genet Res Crop Evol* 1998, **45**:9-17.
9. Gaur PM, Slinkard AE: **Inheritance and linkage of isozyme codominant genes in chickpea.** *J Hered* 1990, **81**:455-461.
10. Gaur PM, Slinkard AE: **Genetic control and linkage relations of additional isozyme markers in chickpea.** *Theor Appl Genet* 1990, **80**:648-656.
11. Kazan K, Muehlbauer FJ, Weeden NF, Ladizinsky G: **Inheritance and linkage relationships of morphological and isozyme loci in chickpea.** *Theor Appl Genet* 1993, **86**:417-426.
12. Labdi M, Robertson LD, Singh KB, Charrier A: **Genetic diversity and phylogenetic relationships among the annual *Cicer* species as revealed by isozyme polymorphism.** *Euphytica* 1996, **88**:181-188.
13. Udupa SM, Sharma A, Sharma RP, Pai RA: **Narrow genetic variability in *Cicer arietinum* L. as revealed by RFLP analysis.** *J Plant Biochem Biotech* 1993, **2**:84-86.
14. van Rheenen HA, Reddy MV, Kumar J, Haware MP: **Breeding for resistance to soil-borne diseases in chickpea.** In *Disease Resistance Breeding in Chickpea* Edited by: Singh KB, Saxena MC. ICARDA, Aleppo, Syria; 1992:55-70.
15. Iruela M, Rubio J, Cubero JJ, Gil J, Millán T: **Phylogenetic analysis in the genus *Cicer* and cultivated chickpea using RAPD and ISSR markers.** *Theor Appl Genet* 2002, **104**:643-651.
16. Sudupak M, Akkaya M, Kence A: **Analysis of genetic relationships among perennial and annual *Cicer* species growing in Turkey using RAPD markers.** *Theor Appl Genet* 2002, **105**:1220-1228.
17. Nguyen TT, Taylor PVJ, Reeden RJ, Ford R: **Genetic diversity estimates in *Cicer* using AFLP analysis.** *Plant Breeding* 2004, **123**:173-179.
18. Mayer MS, Tullu A, Simon CJ, Kumar J, Kraft JM, Muehlbauer FJ: **Development of a DNA marker for fusarium wilt resistance in chickpea.** *Crop Sci* 1997, **37**:1625-1629.
19. Hüttel B, Winter P, Weising K, Choumane W, Weigand F: **Sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.).** *Genome* 1999, **42**:210-217.
20. Lichtenzweig J, Scheuring C, Dodge J, Abbo S, Zhang H-B: **Construction of BAC and BIBAC libraries and their applications for generation of SSR markers for genome analysis of chickpea, *Cicer arietinum* L.** *Theor Appl Genet* 2005, **110**:492-510.
21. Niroj SK, Bhumiika S, Bhatia S: **Isolation and characterization of a sequence-tagged microsatellite site markers in chickpea (*Cicer arietinum* L.).** *Mol Ecol Notes* 2003, **3**:428-430.
22. Winter P, Pfaff T, Udupa SM, Hüttel B, Sharma PC, Sahi S, Arreguin-Espinoza R, Weigand F, Muehlbauer FJ, Kahl G: **Characterisation and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome.** *Mol Gen Genet* 1999, **262**:90-101.
23. Hüttel B, Santra D, Muehlbauer F, Kahl G: **Resistance gene analogues of chickpea (*Cicer arietinum* L.): isolation, genetic mapping and association with a Fusarium resistance gene cluster.** *Theor Appl Genet* 2002, **105**:479-490.
24. Benko-Iseppon AM, Winter P, Hüttel B, Staginnus CM, Muehlbauer FJ, Kahl G: **Molecular markers closely linked to fusarium resistance genes in chickpea show significant alignment to pathogenesis-related genes located on Arabidopsis chromosomes I and 5.** *Theor Appl Genet* 2003, **107**:379-386.
25. Pfaff T, Kahl G: **Mapping of gene-specific markers on the genetic map of chickpea (*Cicer arietinum* L.).** *Mol Genet Genomics* 2003, **269**:243-251.
26. Udupa SM, Robertson LD, Weigand F, Baum M, Kahl G: **Allelic variation at (TAA)_n microsatellite loci in a world collection of**

- chickpea (*Cicer arietinum* L.) germplasm. *Mol Gen Genet* 1999, **261**:354-63.
27. Tekeoglu M, Rajesh N, Muehlbauer FJ: **Integration of sequence tagged microsatellite sites to the chickpea genetic map.** *Theor Appl Genet* 2002, **105**:847-854.
 28. Udupa SM, Baum M: **Genetic dissection of pathotype-specific resistance to ascochyta blight disease in chickpea (*Cicer arietinum* L.) using microsatellite markers.** *Theor Appl Genet* 2003, **106**:1196-1202.
 29. Winter P, Benko-Iseppon AM, Hüttel B, Ratnaparkhe M, Tullu A, Sonnante G, Pfaff T, Tekeoglu M, Santra D, Sant VJ, Rajesh PN, Kahl G, Muehlbauer FJ: **A linkage map of the chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *C. arietinum* × *C. reticulatum* cross: localization of resistance genes for fusarium wilt races 4 and 5.** *Theor Appl Genet* 2000, **101**:1155-1163.
 30. Doyle JJ, Doyle JL, Harbison C: **Chloroplast-expressed glutamine synthetase in *Glycine* and related Leguminosae: Phylogeny, gene duplication and ancient polyploidy.** *Systematic Botany* 2003, **28**:567-577.
 31. Kajita T, Ohashi H, Tateishi Y, Bailey CD, Doyle JJ: **rrcL and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and Allies.** *Systematic Botany* 2001, **26**:515-536.
 32. Romo S, Labrador E, Dopico B: **Water stress-regulated gene expression in *Cicer arietinum* seedlings and plants.** *Plant Physiol Biochem* 2004, **39**:1017-1026.
 33. Sinclair TR, Muchow RC: **System analysis of plant traits to increase grain yield on limited water supplies.** *Agron J* 2001, **93**:263-270.
 34. Colmenero-Flores JM, Moreno LP, Smith CE, Covarrubias AA: **Pvlea-18, a member of a new late-embryogenesis-abundant protein family that accumulates during water stress and in the growing regions of well-irrigated bean seedlings.** *Plant Physiol* 1999, **120**:93-103.
 35. Saxena NP, Krishnamurthy L, Johansen C: **Registration of a drought resistant chickpea germplasm.** *Crop Sci* 1993, **33**:1424.
 36. Saxena NP: **Management of drought in chickpea – a holistic approach.** In *Management of Agricultural drought – Agronomic and Genetic Options* Edited by: Saxena NP. Oxford & IBH Publishers Co Pvt. Ltd, New Delhi, India; 2003:103-122.
 37. Glick B, Pasternak J: *Molecular biotechnology: Principles and applications of recombinant DNA* American Society for Microbiology; Washington DC, USA; 1998.
 38. **ICRISAT chickpea EST database:** [<http://www.icrisat.org/gt1/cpest/home.asp>]
 39. Quackenbush J, Liang F, Holt I, Pertea G, Upton J: **The TIGR Gene Indices: reconstruction and representation of expressed gene sequences.** *Nucl Acids Res* 2000, **28**:141-145.
 40. Ronning CM, Stegalkina SS, Ascenzi RA, Bougri O, Hart AL, Utterbach TR, Vanaken SE, Riedmuller SB, White JA, Cho J, Pertea GM, Lee Y, Karamycheva S, Sultana R, Tsai J, Quackenbush J, Griffiths HM, Restrepo S, Smart CD, Fry WE, Van Der Hoeven R, Tanksley S, Zhang P, Jin H, Yamamoto ML, Baker BJ, Buell CR: **Comparative analysis of potato expressed sequence tag libraries.** *Plant Physiol* 2003, **131**:419-429.
 41. White JA, Todd J, Newman T, Focks N, Girke T, Ilarduya OM, Jaworski JG, Ohlrogge JB, Benning C: **A new set of *Arabidopsis* Expressed Sequence Tags from developing seeds. The metabolic pathway from carbohydrates to seed oil.** *Plant Physiol* 2000, **124**:1582-1594.
 42. Rounsley SD, Glodek A, Sutton G, Adams MD, Somerville CR, Venter JC, Kerlavage AR: **The construction of *Arabidopsis* expressed sequence tag assemblies. A new resource to facilitate gene identification.** *Plant Physiol* 1996, **112**:1177-1183.
 43. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: A new generation of protein database search programs.** *Nucl Acids Res* 1997, **25**:3389-3402.
 44. Coram T, Pang E: **Ascochyta blight of chickpea – searching for resistance gene analogues and characterizing an enriched EST library.** *Proceedings of the 12th Australian Plant Breeding Conference, 15–20 September, Perth, Australia* 2002:17-22.
 45. Beaulieu M, Larson GP, Geller L, Flanagan SD, Krontiris TG: **PCR candidate region mismatch scanning: adaption to quantitative, high-throughput genotyping.** *Nucleic Acid Res* 2001, **29**:1114-1124.
 46. Cobb BD, Clarkson JM: **A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods.** *Nucl Acids Res* 1994, **22**:3801-3805.
 47. Taguchi G: *Introduction to Quality Engineering*, Asian Productivity Organization American Supplier Institute Inc., Dearborn, MI; 1986.
 48. Dreher K, Khairallah M, Ribaut JM, Morris M: **Money matters (I): costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT.** *Mol Breeding* 2003, **11**:221-234.
 49. Ragot M, Hoisington DA: **Molecular markers for plant breeding: comparisons of RFLP and RAPD genotyping costs.** *Theor Appl Genet* 1993, **86**:975-984.
 50. Mace ES, Buhariwalla HK, Crouch JH: **A high throughput DNA extraction protocol for tropical molecular breeding programs.** *Plant Mol Biol Rep* 2004, **21**:459a-459h.
 51. Kresovich S, Luongo AJ, Schloss SJ: **Mining the Gold: Finding Allelic variants for improved crop conservation and use.** In *Managing Plant Genetic Diversity* Edited by: Engels JMM, Ramanatha Rao V, Brown AHD, Jackson MT. IPGRI; 2002:379-386.
 52. Kazan K, Muehlbauer FJ: **Allozyme variation and phylogeny in annual species of *Cicer* (Leguminosae).** *Plant Syst Evol* 1991, **175**:11-21.
 53. Ahmad F: **Random amplified polymorphic DNA (RAPD) analysis reveals genetic relationships among the annual *Cicer* species.** *Theor Appl Genet* 1999, **98**:657-663.
 54. Choumane W, Winter P, Viegand F, Kahl G: **Conservation and variability of sequence-tagged microsatellite sites (STMSs) from chickpea (*Cicer arietinum* L.) within the genus *Cicer*.** *Theor Appl Genet* 2000, **101**:269-278.
 55. Sudupak M: **Inter and Intra-species Inter Simple sequence Repeat (ISSR) variation in the genus *Cicer*.** *Euphytica* 2004, **135**:229-238.
 56. Ladizinsky G, Adler A: **Genetic relationships among the annual species of *Cicer* L.** *Theor Appl Genet* 1976, **48**:197-203.
 57. Ahmad F, Slinkard AE: **Genetic relationships in the genus *Cicer* L. as revealed by polyacrylamide gel electrophoresis of seed storage proteins.** *Theor Appl Genet* 1992, **84**:688-692.
 58. Buhariwalla HK, Gaur PM, Pande S, Crouch JH: **Asian Development Bank.** In *Half-yearly progress report in chickpea for project 'Rapid crop improvement for poor farmers in the semi-arid tropics'* ICRISAT, India; 2002.
 59. Gargav S, Gaur PM: **Phylogenetic relationships of perennial chickpea species *Cicer microphyllum* (Benth.) with its annual relatives as revealed by allozyme polymorphism.** *Indian Journal of Genetics* 2001, **61**:218-222.
 60. Heslop-Harrison JS: **Exploiting novel germplasm.** *Aust J Agric Res* 2002, **53**:873-879.
 61. **RepeatMasker software** [<http://repeatmasker.genome.washington.edu>]
 62. **TIGR BLASTn and tBLASTx analysis** [<http://tigrblast.tigr.org>]
 63. **NCBI BLASTn analysis:** [<http://www.ncbi.nlm.nih.gov>]
 64. **Transeq software:** [<http://www.ebi.ac.uk/emboss/transeq>]
 65. Giegerich R, Meyer F, Schleiermacher C: **GeneFisher – Software Support for the Detection of Postulated Genes.** In *Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology AAAI Press, USA; 1996*:68-77.
 66. **Netprimer analysis** [<http://www.premierbiosoft.com>]
 67. Tegelstrom H: **Detection of mitochondrial DNA fragments in molecular genetic analysis of populations.** In *A Practical Approach* Edited by: Hoelzel AR. IRL Press, Oxford; 1992:89-114.
 68. **Polymorphic information content calculator** [<http://www.agri.huji.ac.il/~weller/Hayim/parent/PIC>]
 69. Rohlf FJ: *NTSYS-pc. Numerical taxonomy and multivariate analysis system, Version 2.0 Applied Biostatistics*, New York; 1998.
 70. Lynch M: **The similarity index and DNA fingerprinting.** *Mol Biol Evol* 1990, **7**:478-484.
 71. Nei M, Li WH: **Mathematical model for studying genetic variation in terms of restriction endonucleases.** *Proc Natl Acad Sci* 1979, **74**:5267-5273.
 72. Felsenstein J: *PHYLP (Phylogeny Inference Package)*, version 3.63 the author, Univ. of Washington, Seattle, Washington; 1995.