SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the Coffea genus

Valérie Poncet, Myriam Rondeau, Christine Tranchant, Anne Cayrel, Serge Hamon, Alexandre De Kochko, Perla Hamon

SSRminingincoffeetreeESTdatabases: potentialuseofEST–SSRsasmarkersforthe*Coffea*genus

ValériePoncet·MyriamRondeau·ChristineTranchant·AnneCayrel·SergeHamon·AlexandredeKochko·PerlaHamon

AbstractExpressedsequencetags(ESTs)from*Coffea canephora*leavesandfruitswereusedtosearchfortypesandfrequenciesofsimplesequencerepeats(EST–SSRs)withamotiflengthof1–6bp.Fromanon-redundant(NR)ESTsetof5,534potentialunigenes,6.8%SSR-containingsequenceswereidentified,withanaveragedensityofoneSSRevery7.73kbofESTsequences.Trinucleotiderepeatswerefoundtobethemostabundant(34.34%),followedbydi-(25.75%)andhexa-nucleotide(22.04%)motifs.Thedevelopmentofuniquegenic SSRmarkerswasoptimizedbyacomputationalapproachwhichallowedustoeliminateredundancyintheoriginalESTsetandalsototestthespecificityofeachpairofdesignedprim-ers.Twenty-fiveEST–SSRsweredevelopedandusedtoevaluatecross-speciestransferabilityinthe*Coffea* genus.Theorthologywassupportedbytheampliconsequencesimilarityandtheamplificationpatterns.The >94%identityofflankingsequencesrevealedhighsequence conservationacrossthe*Coffea* genus.Ahighlevelofpolymorphiclociwasobtainedregardlessofthespeciesconsidered(from75%for *C.liberica*to86%for*C.canephora*).Moreover,thepolymorphismrevealedbyEST–SSRwassimilartothatexposedbygenomicSSR.Itisconcludedthat*Coffea* ESTsareavaluableresourceformicrosatellitemining.EST-SSRmarkersdevelopedfrom*C.canephora*sequencescanbeeasilytransferredtootherCoffea speciesforwhichverylittlemolecularinformationisavailable.Theyconstituteasetofconservedorthologousmarkers,whichwouldbeidealforassessinggeneticdiversityincoffeetrees aswellasforcross-referencingtranscribedsequencesincomparativegenomics studies.

Keywords SSRmining·EST–SSR·*Coffea* ·Transferability· Geneticvariation

V.Poncet(&)·M.Rondeau·C.Tranchant·A.Cayrel· S.Hamon·A.deKochko·P.Hamon
UMR1097DiversitéetGénomesdesPlantesCultivées(DGPC),IRD,InstitutdeRecherchepourleDéveloppement,911 avenueAgropolis,BP64501, 34394MontpellierCedex 5,Francee-mail:poncet@mpl.ird.fr
URL:http://www.dgpc.org/index.html

Introduction

Microsatellitemarkersderivedfromanonymousgenomicsequenceshavebeenextensivelyusedoverthelastdecadebecauseoftheirhighlyinterestingproperties(Ellegren 2004).Thesegenerallyco-dominantandlocus-speciWcmarkershaverevealedhighpolymorphismlevels.Goodtransferabilitybetweenspeciesandsometimesbetweengenerahasoftenbeenreported(Dirlewangeretal.2002;Ralloetal.2003;Gonzalez-Martinezetal.2004;Poncetetal.2004).Thesefeatureshaveproventobeofgreatinterestforgeneticdiversitystudies,geneticandcomparativemapping(WuandTanksley1993; Gonzaloetal.2005;Sahaetal.2005;Vigourouxetal.2005). However,despitetheseadvantages,theyoftencorrespondtonon-codingsequencesandthuscannothelpinseekingcandidate genes.

Expressedsequencetags(ESTs)aresequencedportionsof complementaryDNAcopiesof mRNA—theyrepresentpartofthetranscribedportionofthegenome ingivenconditions.Asexpected,theymainlycorrespondtorelativelyconservedsequences.Techniqueshave beendevelopedtorevealpolymorphismassociatedwithsuchsequences(Catoetal.2001),andstudies basedonintronpolymorphism(lessconservedgenicregions)havealsobeencarriedout(LemandLallemand2003).However,ESTsequenceanalysesrevealedappro

ximately1.5–7.5%ofsequencescontainingmicrosatellitemotifsincereals(Kantetyetal.2002;Thieletal.2003).Amongdicotyledonousspecies,thefrequencyofESTscontainingSSRswasfound torangefrom2.65to16.82%(KumpatlaandMukhopadhyay2005).Therefore,regardlessoftheplantconsidered,aneverincreasingnumberofESTsequencesprovidesacomplementarysourceformicrosatellitemarkeridentification.Although theconservednatureofcodingsequencesmaylimittheirpolymorphism,itshould facilitatecrossamplificationoflociamongphylogeneticallyrelatedspecies(Scottetal.2000)andevengenera.Moreover,theyhaveahighprobabilityofbeingassociated withfunctionalportionsofthegenome.Amongtheirmanyapplications,theseEST-derivedmarkers(EST–SSR)canbeusedtocross-referencegenesbetweenspeciesforenhancingtheresolutionincomparativegenomicsstudiesandidentifyingconservedgenomicregionsamongspeciesandgenera(Brownetal.2001;Decroocqetal.2003;Guptaetal.2003;Sahaetal.2004;Yuetal.2004;Parketal.2005;Varshneyetal.2005b).

Inthe*Coffea* genus,SSRmarkershaverecentlybecomemoreavailable throughtheconstructionofenrichedgenomicmicrosatellitelibraries(Rovellietal.2000;Dufouretal.2001;Baruahetal.2003).Microsatellitemarkerswerefurtherevaluatedforamplificationamong*Coffea* species(Baruahetal.2003;MoncadaandMcCouch2004;Poncetetal.2004).Goodcross-speciestransferabilityandhighgeneticdiversityweregenerallyobserved.However,onlyahandfuloflinkagemapshavebeenconstructed,includingthesemarkers(Lashermesetal.2001;Coulibalyetal.2003;N'Diayeetal.2006).These mapsweremainlyconstructedusingAFLPmarkersbuttheirrelevanceforcoffeetreebreedingprogramsislimited.Moreover,current*Coffea* mapsderivedfromdifferent*Coffea* specieshardlyeversharean adequate numberof common (anchor)mark-erstobeabletobridgemaps.

Publicaccessibilityto*Coffea* ESTdatabasesislimitedtoanSSHlibraryof527non-redundant(NR)ESTsequencesassociated with reactionsto the rustfungus(Fernandezetal.2004).Theseresourceswerevery recentlyenhancedbythegenerationof13,175unigenesfrom*C.canephora*(Linetal.2005).OnlynineEST–SSR markershavebeendevelopedtodate(Bhatetal.2005).

Werecentlydevelopedtwo*C.canephora*ESTsequencesetsinourlaboratoryfromcDNAisolatedfromleavesand fruitsatdifferentdevelopmentand maturationstages(totalof10,420sequences,unpublisheddata).ThefirstobjectiveofthepresentstudywastoidentifyandcharacterizemicrosatellitespresentintheNRsetofourESTstoevaluateitspotentialasasourceformarkerdevelopment.Thesecondaimwastodevelopasetofhighlypolymorphicmarkersthatcouldcross-amplifyanddistinguishbetween*Coffea* species.

Herewereportthedevelopmentof EST–SSRmarkersbasedon*C.canephora*sequencesandtheircross-speciestransferabilitytosix*Coffea* species.Locusorthologywasmonitoredbyanalyzingamplification patternsandbysequencingsomeamplicons.Relevantinformationaboutthemutationmodelandtheevolutionoftheselociwasalsonoted.Polymorphismsdetectedwithinandbetweenasetof*Coffea* specieswasalsoanalyzedtoassesswhetherthesemarkerscouldbeusefulfordiversitystudiesanddistinguishingbetween*Coffea* species.Weidentifiedasetofanchormarkers,mostofthemwithingenesoffunctionalrelevance,whichwouldbehelpfulfor functionalandcomparativemappingwithinthe*Coffea* genus.

Materialsandmethods

PlantDNA

CoVee treesweremaintainedin a tropical greenhouseattheIRDresearchcenterinMontpellier (France).

TotalDNAfromfullydevelopedleaveswasextractedaccordingtoKyetal.(2000).Sevenspecies(*C.canephora* Pierre,knownasRobusta (CAN), *C.eugenioides*Moore(EUG),*C.heterocalyx*StoVelen(HET),*C.liberica*Portères(LIB),*C.dewevrei*Portères(DEW),*C.*sp.Moloundou(MOL),and *C.pseudozanguebariae*Bridson(PSE))representingthethreeAfricanmaingeographicalclades(Lashermesetal.1997)wereanalyzed.Polymorphismwasassessedin12CAN,10DEW,10PSEand8LIB,representativeofthegeneticdiversityofthe fourspecies,plus2EUG,2MOL,andtheonlyknownHETindividual.

DataminingforSSRmarkers

ThecoffeeESTdatabasesusedinthisprojectweredevelopedinourlaboratory(deKochkoetal.,unpublisheddata).Theycontained10,420sequencesderivedfromfruit(5,814sequences)andleaf(4,606)cDNAlibraries(validsequencessubmittedtoGenBankunderaccessionnumbersEE191792–EE200565).Therawchromatogramswereprocessedusing ESTdbsoftware

(http://www.mpl.ird.fr/bioinfo/).ESTdbisasetofanalyticalproceduresthatautomaticallyverify,cleanse,storeandanalyzeESTsgeneratedinourlaboratory.ESTdbhasthreemajorcomponents:(1)apipelineanalysisprogrambasedonpublicsoftwareintegratedbyanin-housedevelopedscript,(2)theresultsandrelatedinformationarestoredinarelationaldatabaseaccessiblethrough(3)awebinterface(Fig.1).ThefunctionsoftheESTsequenceswerepredictedthroughsimilaritysearchesfromproteinsequenceGenBankdatabases(http://www.ncbi.nlm.nih.gov)usingBLASTx(Altschuletal.1997).Potentialunigenes(contigsandsingletonsfromallESTsequencesprocessedsimultaneously)wereidentifiedafter clustering for aNRESTset.

Theseunigenesequenceswerescreenedforthepresenceofperfectandimperfectmicrosatellitesusing SSRpipeline,aPerlscriptdevelopedbyDubreuil-Tranchant(SSRpipelineispubliclyavailableathttp://www.mpl.ird.fr/bioinfo/).Thisprogramintegratedthree publicsoftware packages:(1)a Perlprogramdevelopedby Cartingour(http://www.gramene.org)whichdetectsperfectmicrosatellites,(2)RepeatTandemFinder, anSSRrepeatWnder(http://www.tan-dem.bu.edu),and (3)Primer3,aPCRprimerdesignprogram(RozenandSkaletsky2000).Moreover,thisprogramallowedustocheckthespecificityofprimerpairsbyblastingagainsttheESTsequences. Thistoolscreenedeachsequenceforssrs.Theparameterswereresetfordetectionofmono-,di-,tri-,tetra-,penta-,andhexanucleotidemotifswithaminimumof15,9,6,5,4,and3repeats,respectively.Themajorprimerdesignparameters weresetasfollows:primerlengthfrom18to21bp(optimum20),PCRproductsizefrom100to300bp,optimumannealingtemperature60°C.

Asetof25EST–SSRprimerswasfurtheranalyzed(sequencessubmittedtoGenBankunderaccessionnumbersDQ778713–DQ778737).Theywereselectedbasedonthefeatureofthesequence:wekeptinpriority

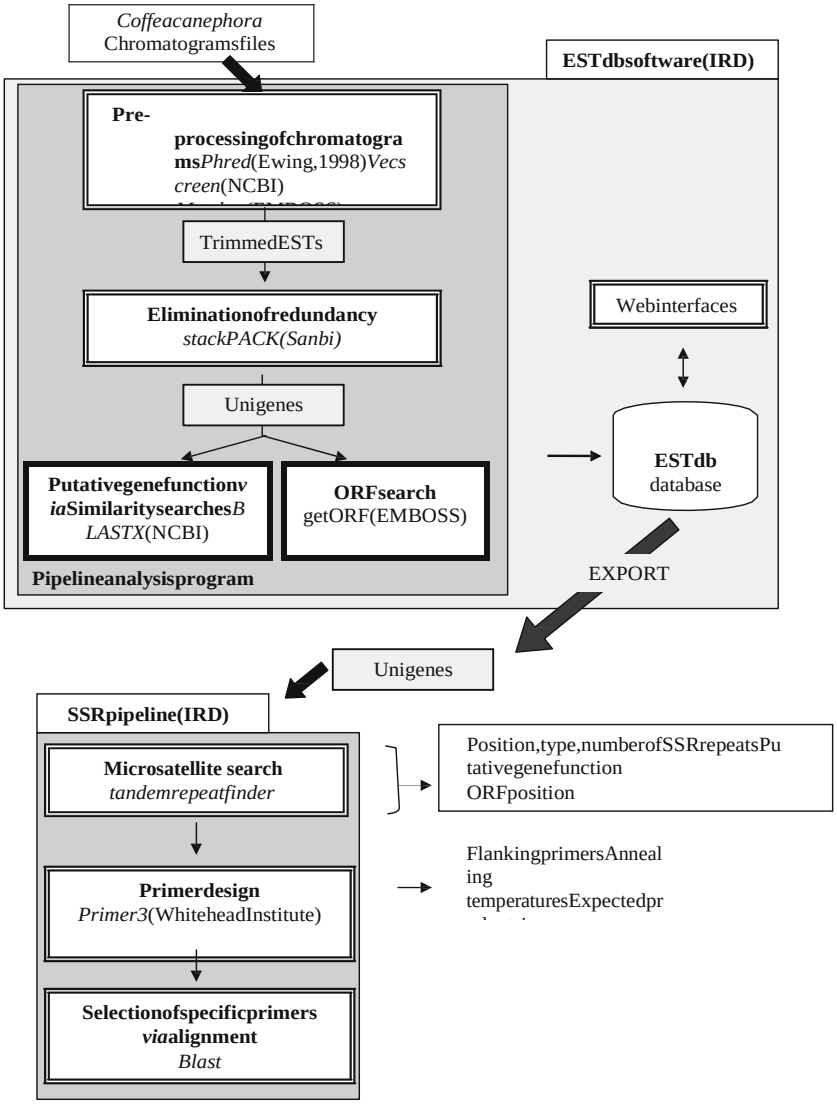

Fig.1 Flowchartof the clusteringandannotationofESTs*via*theIRDpipelineESTdb,dataminingofSSR-containingESTsandprimerdesignforanon-redundantSSRset

singletonESTs.Everyforwardprimerwas5-tailedwiththeM13sequence5-CACGACGTTGTAAAACGAC-3.TheprimersweresynthesizedbyMWG-BiotechAG(Ebersberg,Germany).

Polymerasechainreaction(PCR)amplificationand visualisationofamplicons

Polymerasechainreactionamplificationswerecarriedoutasdescribedin Coulibaly et al. (2003) and Poncet et al. (2004) usingatouchdownPCRprofileoptimisedforeachsetofprimers: touchdown60–55°C ortouch-down55–50°C.PCRproductsweredetectedonanIR^{2}AutomatedDNASequencer(LI-COR,model4200L-2,Lincoln,NE,USA)usinganM13primercoupledtotheinfraredtagIRD700orIRD800andaftermigra-tionon25cm6.5%KBplus(LI-COR,CAT#827-05607)polyacrylamidegels.

Thegelimageswereprocessed bySAGAGT™software(LICORBiotech)toestimatethesizeofampliconsaccordingtoa50–350bpsizestandard(LI-COR,CAT#829-05343,829-05344).

Microsatellitecloningandsequencing

GenomicDNAswereamplifiedwiththeappropriateforwardandreverseprimersbutwithouttheinfraredfluorescentM13.TheproductswerepurifiedusingEZNACycle-Pure(OMEGABio-Tek,Doraville,GA,USA)andclonedontothepCR®4-TopoplasmidusingtheTOPOTAcloningkit(Invitrogen,Groningen,TheNetherlands)accordingtothemanufacturer'sinstruc-tions.

CloningefficiencywascheckedthroughPCRampli-Wcationofseveralcolonies.Theresultingproductswererunonpolyacrylamidegels,withtheinitialgeno-micDNAamplificationoutcomeascontrol.Thereso-lutionusedatthisstepwasespeciallyimportanttodistinguishthetwoallelesfromheterozygousindividu-als.

AfterplasmidDNApurification(Sambrooketal.1989),theclonedPCRfragmentsweresequencedbyMWG-Biotech(http://www.mwg-biotech.com/html/all/index.php).Thesequencedatahavebeensubmitted toGenBankunderaccessionnumbersDQ787368–DQ787384.Sequenceswereeditedandanalyzedusing theDNASTARsoftwarepackage(Lasergene,Madi-son,WI,USA).

Dataanalysis

Numberofalleles,observedheterozygosity(Ho),genediversity(orexpectedheterozygosity,He),polymorphism

informationcontent(PIC),nullallelefrequency(rb)andheterozygotedeficiencyorexcess(Fis)werecalculatedforeachlocusandforthefourspeciesrepresentedbymorethanfivegenotypes,asindicatedinPoncet etal.(2004),usingPowerMarker(LiuandMuse2005).Foreachspecies,polymorphiclociat0.05thresholdfrequencyandmeanallelenumberperpolymorphiclocuswereevaluated.

Results

ESTunigenesets

Atotalof9,820(94%)validESTswereobtainedfromall*Coffea canephora*chromatogramfiles.Theaveragelengthwas602bp.Afterclusteringandassembly,5,534unigeneswereidentified,including3,747singletonsand1,787contigs.TheBLASTresultsrevealedthatabout22.3%oftheseNRESTsshowednosimilaritytoanyGenBanksequences.Lessthan0.03%oftheESTswaspredictedtobefromplastidcDNA.

Frequencyanddistributionof*C.canephora*EST–SSR

EST–SSRswereminedfromaNRESTsetof5,534potentialunigenes.Accordingtothesearchcriteriaadopted,431uniqueSSRswerefoundwithinatotalof376unigenes.Thechanceoffindinganssr-containingsequenceintheNRESTdatabasewasthus6.8%,withanaveragedensityofonemicrosatelliteevery7.73kb.

Trinucleotidemotifswerethemostabundant(34.34%),followedbydi-(25.75%),andhexa-nucleotiderepeats(22.04%)(Fig.2).Themostabundanttrinucleotiderepeatmotifwas AGG/TCC(23%),followedbyAAG/TTC(20.3%)andAAC/TTGandAAT/TTAweretheleastabundantmotifs(3.4%each)(Fig.2).Amongthedinucleotidemotifs,GA/CTwasthemostabundant(62.2%)andnoGC/CGmotifwasfound.Forallrepeatclasses,themeanSSRlengthwas 20.6bp,butwashigherfordinucleotides(24.1bp),with amaximum of94bp.

Outofthe25EST–SSRsusedfordesigningprimers,ninewerefoundintranslatedregions,eightinthe5untranslatedterminalregion(5UTR)andfourinthe3UTR.Thefourothersequencesgeneratednohitsinthesimilaritysearch(Table1).

ConservationofthologousSSRlociamong *Coffea* species

Toassesstheconservationof*C.canephora*microsatellitelociacross*Coffea* species,we testedthe

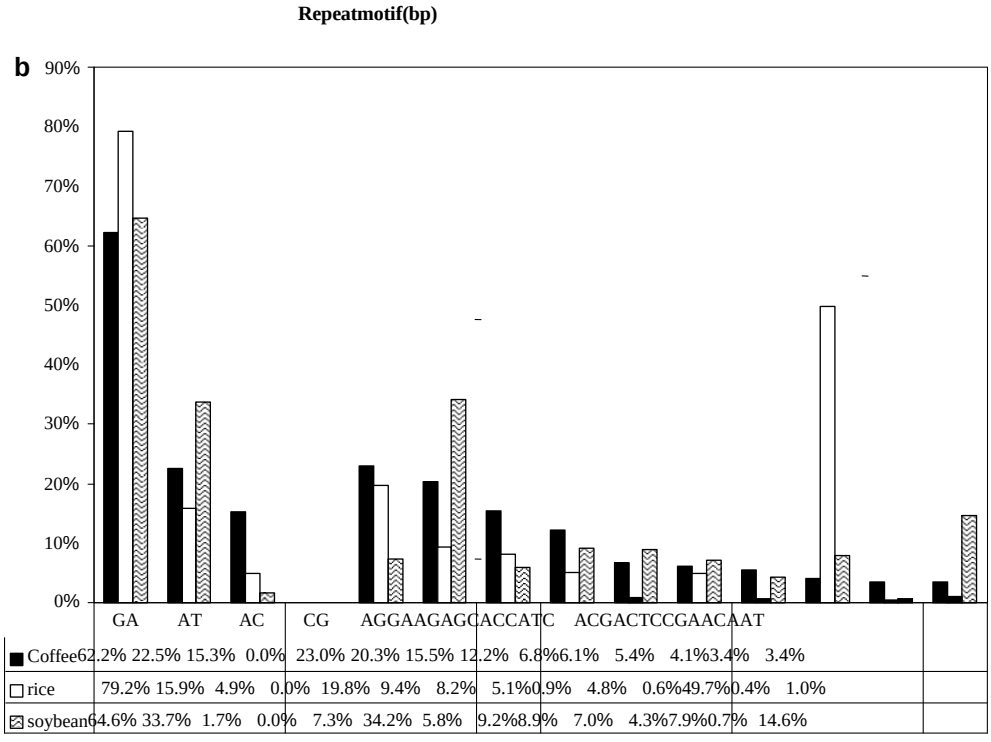

Fig.2 SSRfrequencyintheunigene*Coffea canephora*ESTset(riceandsoybeandatafromGaoetal.2003)**a**accordingtotherepeatmotiflength,**b**accordingtothemotifofdi-andtrinucleotiderepeats

cross-amplificationof 25primerpairsonsevenspeciesrepresentingthreedifferent*C offea* phylogeneticclades.Out of25, 24primerpairs amplified atleastonespe-cies(onlyES21couldnot)(Table2).Twomicrosatel-lites(ES3,ES15)did notamplifywell,and theamplifiedproductswere substantiallylargerthanexpected. Finally,cross-speciestransferabilitywithintheCoffea genuswasveryhighsinceeachprimerpairyieldedaPCRprod uctinanaverageof6.1species(outofseventested)and14(56 %)gaveamplificationwith alltested*Coffea* species.

Oneortwobandspergenotypewereobservedinallbutone ofthereadableamplificationpatternsobtained,suggestingthat therewas selectiveamplification ofonelocus.TheES11primerpairamplified3–6productswithsome*C.liberica*genotypes,indicatingapossible

triplicatesequence—thiswasconsideredasmissingdataforthesubsequentanalyses.

Giventhatprimersweredesignedonthebasisof *C. canephora*ESTsequences, ampliconsproducedfrom*C.canephora*genomicDNAhad the expectedsizeinmostcases(Table2),consistent withlocus-specificamplification.Inallother*Coffea* species,theallelesizesalsodidnotsubstantially differfromtheexpectedsize,whichwasconsistentwithselectiveorthologousamplification.The microsatelliteallelesizedistribu-tions, expressedinbase pairs,fittedthe stepwisemodelinmostcases(Table 2)—withallelicdifferences,whicharemultiplesoftherepeatmotifevenifsomeintermediateallelicstatesweremissing.Some“jumps”inallelesizedistribution,associatedwithmixed distribu-tionpatterns, mightbeattributed to theadditional

Table1
Expressedsequencetag(EST),putativegenefunctionbasedontBLASTxsearchofGenBanksequences(analysisonDecember2005),SSRpositionin thecDNAsequence(correspondingrepeateda minoacidwhentranslated)

Coffea EST–SSR				Homologous sequences				
Locus	Repeatmotif	Expected ampliWcon size(bp)	SSRposition	GenBank accession No.	Species	BLAST *E*value	Identity (%)	Putativefunction
ES1	$(GCA)_8$	308	ORF(Q)	AAD02552	*Petuniahybrida*	2e¡06	64	PGPS(P-glycoprotein)
ES2	$(A)_{14}(CA)_{19}$	349						Nohit
ES3	$(TTTCT)_9$	371	5 UTR	CAB75429	*Nicotiana plumbaginifolia*	1e¡41	90	Oligouridylate bindingprotein
ES4	$(CT)_{14}$	337	3 UTR	CAA64545	*Hordeumvulgare*	4e¡21	73	Xyloseisomerase
ES5	$(TTC)_{24}$	309	ORF (S)	AAL65125	Glycinemax	2e¡31	66	GT-2transcriptionfactor
ES6	$(CCA)_2(CCT)_{11}$	228	5 UTR	AAN71762	*Solanumtuberosum*	1e¡41	75	CinnamoylCoAreductase2
ES7	$(TCA)_{18}$	298						Nohit
ES8	$(TC)_{26}$	323	5'UTR	AAB84350	*Arabidopsisthaliana*	3e¡31	72	Heatshocktranscriptionfactor
ES9	$(TTC)_{10}$	205	5 UTR	AAD55726	*Vitisriparia*	7e¡74	75	Galactinolsynthase
ES10	$(ACC)_{11}/(TCC)_6$	403	ORF(T/S)	AAP40485	*Arabidopsisthaliana*	7e-47	56	AP2domaintranscriptionfactor
ES11	$(TTC)_{10}$	240	5 UTR	AAN28269	*Gossypiumhirsutum*	1e¡63	94	myb-Liketranscriptionfactor
ES12	$(AAT)_4(CAG)_{13}$	156	ORF(Q)	AAM54033	*Populustremula*	8e¡69	67	PIN1-likeauxintransportprotein
ES13	$(AG)_{16}$	235	5 UTR	AAW28999	*Antirrhinummajus*	3e¡44	53	1-Deoxy-D-xylulose-5-phosphate synthase
ES14	$(CCA)_5/(CCA)_{10}$	289	ORF(T)	AAF01534	*Arabidopsisthaliana*	1e¡53	56	Proteinkinase
ES15	$(CT)_{12}$	308	5 UTR	AAD42941	*Catharanthusroseus*	2e¡26	81	Ubiquitin-conjugatingenzymeE2
ES16	$(TA)_{15}$	186	3 UTR	AAC62396	*Ricinuscommunis*	2e¡54	72	Cysteineendopeptidaseprecursor
ES17	$(TG)_{16}$	300	5 UTR	CAG18177	*Arabidopsisthaliana*	1e¡74	83	UDP-galactosetransporter
ES18	$(GGATCA)_5$	202	ORF(SG)	AAM67530	*Arabidopsisthaliana*	1e¡64	58	Unknownprotein
ES19	$(CA)_{13}(TA)_6$	259	ORF(TH)	AAQ14193	*Solanumtuberosum*	2e¡62	67	Calciumhomeostasisregulator
ES20	$(AAC)_9$	197	ORF(N)	AAO45753	*Cucumismelo*	1e¡51	57	RING/C3HC4/PHD zincWngerprotein
ES21	$(T)_{28}$	254	3 UTR	AAO33591.	*Arachishypogaea*	7e¡50	65	Earlylightinducedprotein
ES22	$(TTTTTTC)_4$	207	3 UTR	P25317	*Nicotianatabacum*	7e¡50	65	glutathione*S*-transferase (auxin-regulatedprotein)
ES23	$(AG)_{16}$	230						Nohit
ES24	$(TCTCC)_7$	224						Nohit
ES25	$(CCT)_9$	207	ORF(P)	AAG45420	*Chlamydomonasreinhardtii*	6e¡10	33	Vegetativecell wall proteingp1

Table2 Numberandmolecularsizerangesof25EST–SSRtestedoneight*Coffea* species:*C.canephora*CAN,*C.dewevrei*DEW,*C.pseudozanguebariae*PSE,*C.liberica*LIB, *C. heterocalyx*HET,*C.eugenioides*EUG,and*C.*sp.MoloundouMOL(samplesizegiveninbrackets)

EST–SSR	Expecte dsize(bp)	Allelesizerange[a](bp)/alleleno.							Totalno ofalleles	Typeofallelesizedi stribution
		CAN(12)	DEW(10)	PSE(10)	LIB(8)	HET(1)	EUG(2)	MOL(2)		
ES1	327	318–327	306–318	306–321	306–318	318	318	327		Stepwise
		3	4	2	3	1	2	1	7	
ES2	368	356–371	356–363	342–356	342–358	342	346	349		Stepwise
		6	4	3	4	1	1	1	10	
ES3	390	>2,000	>2,000	>2,000	>2,000	>2,000	>2,000	>2,000		–
		–	–	–	–	–	–	–		
ES4	356	344–372	338–396	338–362	346–372	350	362–368	–		Mixed
		8	8	5	6	1	3	–	15	
ES5	328	309–315	291–306	300–309	297–315	294	291–294	300		Stepwise
		3	5	4	4	1	2	1	9	
ES6	247	256	238–268	244–259	256–298	238–256	238–256	262–259		Mixed
		1	4	3	4	2	3	2	9	
ES7	317	300–324	302–315	288–290	288–307	307	288–302	332		Nearlycontinuous
		10	6	2	6	1	3	1	19	
ES8	251	236–248	246–258	222–234	236–258	228	226–234	224		Stepwise
		5	4	3	5	1	2	1	13	
ES9	224	204–230	210–218	214	204–214	–	215	–		Nearlystepwise
		6	3	1	4	–	1	–	9	
ES10	422	395–415	392–420	404	389–404	–	–	–		Nearlystepwise
		7	5	1	5	–	–	–	11	
ES11	259	246–266	240–269	254–260	240–266	246	234	243–263		Nearlystepwise
		8	6	2	9	1	2	2	14	
ES12	175	160–178	169–178	151–166	169–178	154	160–163	169		Stepwise
		5	3	2	4	1	2	1	9	
ES13	254	235–255	237–241	233–237	233–241	237	239–241	–		Stepwise
		10	2	2	2	1	2	–	12	
ES14	308	297	291–297	294–297	291–294	297	–	297		Stepwise
		1	3	2	2	1	–	1	3	
ES15	327				Nonscorable					–
ES16	205	180–222	206–212	195–212	207–236	–	196–198	196		Mixed
		11	4	2	8	–	2	1	17	
ES17	319	»640	»640	»640	»640	»640	»640	»640		
ES18	221	218–222	212	206–218	212–218	206	206	206–218		Stepwise
		2	1	2	2	1	1	1	4	
ES19	278	256–284	250	250–252	244	252	252	242		Stepwise
		9	1	2	1	1	1	1	13	
ES20	216	204–219	212	215–229	209–212	212	212–219	212		Mixed
		4	1	7	2	1	2	1	11	
ES21	273	–	–	–	–	–	–	–	0	–
ES22	226	213–232	–	–	213–220	–	230	–		Mixed
		6	–	–	2	–	1	–	6	
ES23	249	236–262	238–240	240–270	236–260	246	236	244–246		Stepwise
		8	2	12	5	1	1	2	16	

Table2continued

EST–SSR	Expected size(bp)	Allelesizerange[a](bp)/alleleno.							Totalno ofalleles	Typeofallelesize distribution
		CAN(12)	DEW(10)	PSE(10)	LIB(8)	HET(1)	EUG(2)	MOL(2)		
ES24	243	228–247	225–230	230–240	225–235	242	235	230		Mixed
		4	2	3	2	1	1	1	7	
ES25	226	224–227	218–221	230–239	218–221	230	239	–		Stepwise
		2	2	2	2	1	1	–	6	
No.markers		21	20	20	20[c]	17	19	15	21	
Totalalleleno.Averag		119	70	62	73[c]	18	33	18	220	
alleleno.Polymorphi		5.7	3.5	3.1	3.7[c]	1.1	1.7	1.2	10.5	
cloci[b]Averagealleleno		18	16	16	15[c]					
./polymorphicloci		6.4	4.1	3.5	4.1[c]					

[a]M13tailincludedinthesize

[b]Polymorphicloci atthe5%levelforsamplesizeequaloroverWeindividuals

[c]ES11locusnotincluded(multilocusmarkerinLIB)

contributionofinsertions/deletions(indels)intheamplifie dregion.Thispatternwasobservedbetweenspeciesatthe ES16locus.However,somesubstantialallelesizedifferen ceswereobservedwiththreeprimerpairs(ES3,ES15andE S17),withampliconsmuchlargerthanexpected,irrespect iveofthespecies.Forinstance,theobservedallelesizeforE S17wasabout640bpforallspeciesversus319bpexpected.

Productsobtainedbyamplificationwith ES2,ES4,ES6,E S16,ES17,ES19,andES25primerpairswereclonedandsequ encedtoconfirmtheorthology.Theywerealsoanalyzedtoide ntifytheoriginofwithin-andbetween-speciespolymorphismas wellasdiscrepanciesbetweentheobservedandexpectedallel esizes(Table3).Thegenotypeschosenforcloningpresented allelesizesattheextremesofthedistributionofagivenmarker. Therelativesequencesizeswereinagreementwiththesizese valuatedaftermigration,withameanSAGAsoftwareunderestimationof ¡6.72bp.ThegenomicDNAsequencesclarifiedthreemainpoints:

(1) therewasstrongsequencehomologyallalongtheregions flankingtherepeatmotifbetweenampliconsderivedfromdif ferentspecies(mean96.9%identity),inagreementwiththeamplificationoforthologousloci;(2)theobservedpolymorphismmainlyresultedfromvariationsinrepeatnumber,althoughsomemorecomplexmutationpatterns(ES6,ES16)werealsodetected,involvingadditionalvariationofindelsintheflankingsequences;(3)thelargesizeofES17genomicfragmentswasduetoamplificationofa426bpintroninallspecies,althoughthepolymorphismstillresultedfromadiVerentnumberofrepeats

Within-andbetween-speciespolymorphism

Thenumberandmolecularsizerangesofallelesobtainedwitheach EST-SSRareshowninTable2togetherwiththenumberofpolymorphicloci.Onelocusappearedtobemonomorphicfor*C.liberica*(ES19),twofor*C.canephora*(ES6,ES14)and *C.pseudozanguebariae*(ES9,ES10),and threefor *C.dewevrei*(ES18,ES19,ES20).Fortheremaining loci,upto12allelesperlocuswererecorded(ES23amplification in*C.pseudozanguebariae*).Forthethreespeciesrepresentedbyone(*C.heterocalyx*),two(*C.*sp.Moloundou)orthree(*C.eugenioides*)genotypes,18allelesat17and 15lociweredetectedfor*C.heterocalyx*and*C.sp.*Moloundou, respectively,while33allelesat19lociwereobservedfor*C.eugenioides*.

Whenconsideringmarkerswithdi-,tri-,andmoremicrosatelliterepeatedmotifs,theaveragenumberofallelesdetectedperlocuswas14.3(86alleles,6loci), 9.7(107alleles,11loci)and5.7(17alleles,3loci),

Table3 Allelesequenceanalysesatsevenmicrosatellitelocifordi Veren tspeciesgenotypes(homozygotesorheterozygotes)rela-tivetothe *C.canephora*ESTsequence(*ESTdb*):total length, microsatelliterepeat stretches, introns, indels(insertions/dele-tions),and%similarityoftheXankingsequences

Locus	*Coffea* species	Individual	Cloneno.	Size(bp)	Polymorphism	Percentagesimilarityo fXankingsequences
ES2	CAN	*ESTdb*		349	$(A)_{14}(CA)_{19}$	100
	PSE	P-8056	ES2-11	341	$(A)_{14}(CA)_{15}$	97.3
	CAN	C-IF182	ES2-21	357	$(A)_{18}(CA)_{21}$	98.0
			ES2-22	351	$(A)_{18}(CA)_{18}$	98.3
ES4	CAN	*ESTdb*		337	$(CT)_{14}$	100
	DEW	D5765	ES4-15	349	$(CT)_{20}$	97.4
ES6	CAN	*ESTdb*		228	$(CCA)_2(CCT)_{11}$	100
	PSE	P-55	ES6-12	227	$(CCA)_3(CCT)_8$	95.5
			ES6-11	220	$(CCA)_3(CCT)_6$	95.5
	LIB	L-A12	ES6-25	255	$(CCA)_4(CCT)_{10}$+20bp	93.9
			ES6-26	247	$(CCA)_2(CCT)_{10}$	95.5
ES16	CAN	*ESTdb*		186	$(TA)_{15}$	100
	CAN	C-IF72	ES16-12	199	$(TA)_{21}$	96.1
			ES16-25	165	$(TA)_4$	95.9
	LIB	L-A22	ES16-15	206	$(TA)_{17}$+17bp	95.5
ES17	CAN	*ESTdb*		300	$(TG)_{16}$	100
	PSE	P60	ES17-10	718	$(TG)_{10}$+intron426pb	97.9
ES19	CAN	*ESTdb*		260	$(CA)_{13}(TA)_6(T)_2$	100
	CAN	C-IF461	ES19-15	266	$(CA)_{10}(TA)_{11}(T)_4$	98.6
			ES19-16	246	$(CA)_2(TA)_7(T)_9$	98.2
	LIB	L-A11	ES19-26	233	$(CA)_1(TA)_0(T)_9$	99.1
ES25	CAN	*ESTdb*		207	$(CCT)_9$	100
	DEW	D-5462	ES25-11	200	$(CCT)_6$	98.3
	PSE	P-1	ES25-24	219	$(CCT)_{13}$	96.0

CAN*C.canephora*,DEW*C.dewevrei*,PSE*C.pseudozanguebariae*,LIB*C.liberica*

respectively.Whenconsideringthelocationofthemicrosatel litemotifinrelationwiththegeneannota-tion,anaverageof8.1(73/9),12.7(11/3),and11.4(57/ 5)allelesperlocuswereobtainedforORF,3UTRand5UTRlo cations,respectively.

Clearlyreadablelociwereanalyzedtoevaluategeneticdiver sityparameters,whiletakingintoaccountonlysin-gle-locusmarkers which gave anamplification withmorethanfivegenotypesperspecies,between16(*C.liberic a*)to21(*C.canephora*)(Table4).Ahighlevelofpolymor-phiclociwasobservedwhateverthespeciesconsidered(from75 %(*C.liberica*)to86%(*C.canephora*)).Themeanallelenumbe rperpolymorphiclocuswashighestfor*C.canephora*(6.4)wh ilelowerbutsimilarvalueswereobtainedfor*C.dewevrei*, *C.liberica*, and *C.pseudozanguebariae*(4.1,4.1, and3.5,respectively).The observed heterozygosity was lowest for *C.pseudozanguebariae*(0.28)andhighestfor*C.liberica*and*C .canephora*(0.52and0.51,respectively).ThePICvaluerang edfrom0(monomorphlocus)to0.91(ES23for*C.pseudozang uebariae*,highlypolymorphiclocus)acrossthefourspecies, withthelowestaveragevalueobtainedfor*C.pseudozangueb ariae*(0.40)andsimilaraveragevaluesfor*C.canephora*and*C .liberica* (0.59and0.57).Therewasnosignificantcorrelationbetweeni nterspecificPICvaluesexcept forthe *C.liberica*/*C.dewevrei*combination(0.78;*P*=0.003).

Althoughvariationsbetweenlociwereobserved,aglobal heterozygotedeficit(Fis=0.10–0.26)andthepresenceofnullalleles(rb= 0.03–0.09)couldbeesti-matedforallspecies.Thelowestvalueswereobtainedfor*C.ca neph*oraand *C. liberica*,whilethe highestwereobtainedfor*C.dewev*reiand*C.pseudozangueb ariae*.

DiagnosticallelesandlocusidentiWcation

Allelesizesidentifiedforthecultivatedspecies *C.canephora*(CAN)werecomparedtotheothers.DEW/PSE speciescomparisonwasalsoconsideredsincethiscouplewasi nvolvedinthefirstinterspecificmap(Kyetal.2000).From4to1 6locicouldbeidenti-fiedasdiagnosticmarkers,i.e.markersforwhichtherewereno sharedallelesbetweenspecies(Table4).Thesmallestnumber ofsuchlociwas obtainedfortheCAN/LIBcomparisonandthegreatestforCA N/MOL.Moreover,withinthesemarkers,somewerecharact

er-izedbya non-overlappingallelesizedistributionforbothspecies:fourfor CAN/PSE,sevenforCAN/HET,

Table4 Diversitystatisticsfor21EST–SSRin*Coffea*
speciesasdescribedbytheexpected(He)andobserved(Ho)heterozygosities,thepolymorphisminformationcontent(PIC),nullallelefrequency(rb),andheterozygotedeWciencyorexcess(Fis)

EST–SSR	Heobs							PIC				Nullallele frequency(rb)				Fis				Diagnosticloci speciescombination					
	CAN	DEW	PSE	LIB	HET	EUG	MOL	CAN	DEW	PSE	LIB	CAN	DEW	PSE	LIB	CAN	DEW	PSE	LIB	CP	CH	CL	CE	CM	PD
ES1	0.17	0.75	NA	0.43	01/2	0		0.41	0.73	NA	0.52	0.17	¡0.01	NA	0.06	0.59	¡0.02	NA	0.18						
ES2	0.92	0.6	0.3	0.83	0000.78	0.47		0.470.65	¡0.08	¡0.09	0.11	¡0.11	¡0.19	¡0.29	0.35	¡0.28					++		++	++	
ES4	0.77	0.78	0.5	1	0	2/3	–	0.740.84	0.75	0.81	¡0.02	0.07	0.14	¡0.11	¡0.04	0.16	0.33	¡0.24						*	
ES5	0.17	0.33	0.43	0.83	0000.64	0.58		0.580.64	0.29	0.16	0.1	¡0.12	0.74	0.43	0.26	¡0.3					++		++	++	
ES6	0	0.2	0.1	0.33	1/1	1/2	2/2	0	0.51	0.19	0.69	0			0.21	0.07	0.21	NA*	0.61	0.46	0.52	+				++	+
ES7	0.92	0.5	0.3	0.5	0	1/3	0	0.86	0.55	0.5	0.78	¡0.04	0.030.13		0.16	¡0.08	0.09	0.39	0.36					++	++
ES8	0.77	0.4	0.22	0.57	01/2	0		0.74	0.73	0.2	0.71	¡0.02	0.19	¡0.01	0.08	¡0.04	0.45	¡0.08	0.2		++	++		++	++	++
ES9	0.77	0.25	NA	0.33	–	0/1	–	0.75	0.59	NA	0.65	¡0.01	0.22	NA	0.19	¡0.03	0.58	NA	0.49		+	*		+	*	
ES10	0.25	0.4	NA	0.33	–	–	–	0.780.72		NA	0.78	0.3	0.18	NA	0.25	0.68	0.44	NA	0.57		+	*		*	*	
ES11	0.85	0.6	0.3	trip	0	1/2	1/20.7		0.73	0.46	trip	¡0.09	0.080.11		trip	¡0.21	0.18	0.34	trip					++		
ES12	0.83	0.22	0.1	0.83	01/2	0		0.68	0.63	0.1	0.65	¡0.09	0.25	0		¡0.11	¡0.23	0.65	¡0.05	¡0.28	+	++			+	++
ES13	0.77	0.2	0.5	0.17	03/3	–		0.75	0.48	0.46	0.38	¡0.01	0.19	¡0.03	0.15	¡0.03	0.58	¡0.1	0.56						*	
ES14	0	0		0.11	0.5	0	–	000.46		0.11	0.38	0			0.32	NA	¡0.09	NA*	1		NA*	¡0.33		++	*		
ES16	0.92	0.75	0	0.83	–	1/2	0/1	0.88	0.62	0.28	0.85	¡0.03	¡0.08	0.22	0.01	¡0.05	¡0.22	1			0.02	+	*			+	
ES18	0.1	0		0.3750.2	0	2/3	0	0.1	0		0.43	0.180		0	0.04	¡0.02	¡0.05	NA*	0.12	¡0.11			++		++	++	+
ES19	0.62	NA	0	001/2	0		0.83	NA	0.41	00.12	NA	0.29	0			0.25	NA	1		NA*	++	++	++		++	
ES20	0.1	0		0.89	0.330		0	0/1	0.51	0		0.76	0.44	¡0.07	0		¡0.07	0.08		¡0.2	NA*	¡0.17	0.25			
	++	ES22		0.77	–		–	NA	–	0		–	0.72	–		–	NA	¡0.03	–		–	NA	¡0.07	–	–	NA	*
	*	*																								
ES23	0.46	0.5	0.78	NA	0	01/2		0.84	0.38	0.91	NA	0.2	¡0.09	0.07NA		0.45	¡0.33	0.14	NA			+				
ES24	0.38	0.13	0.33	NA	00/1	0/10.38		0.12	0.29	NA	0				¡0.01	¡0.03	NA	¡0.01	¡0.07	¡0.15	NA			+	+	+	
ES25	0.2	0.38	0.33	NA	00/1	–		0.32	0.43	0.28	NA	0.09	0.04	¡0.04	NA	0.38	0.13	¡0.2	NA		++	++	++	++	*	++
Average	0.51	0.37	0.28	0.52				0.59	0.50	0.40	0.57	0.030.09		0.06	0.04	0.10	0.26	0.23	0.11	Total	9 +	1*8 +	4* 4+	8 + 2*	10 +	6*7+

Diagnosticloci: withnooverlappingallelicrangebetweenspeciesofeachcouple

*NA*notapplicable(lessthanWveindividualsanalysedor divisionbyzero),–indicatesnoamplification,tripdenotestriplicatedlocus,++indicatesnooverlappingallelesizedistributions,+indicatesnocommonallelesbetweenthetwospecies,*indicatesnoampliWcationwiththenon-CANspecies,C:*C.canephora*,P:*C.pseudozanguebariae*,H:*C.heterocalyx*,E:*C.eugenioides*,M:*C.*sp.Moloundou,D:*C.dewevrei*,L:*C.liberica*

threeforCAN/LIB,sixforCAN/EUG,sevenforCAN/MOL, and Wve for PSE/DEWcomparisons.

Discussion

ESTsarearichsourceofSSRsin*Coffea*

Atotalof431uniqueSSRswereidentiWedfrom5,534potentialunigenes.Theseresultsclearlydemonstratedthat*Coffea* ESTsareavaluableresourceforminingSSRmarkers.

RandomsequencingwithincDNAlibrariesleadstoahighproportionofredundantESTs.Inourstudy,redundancywaseliminatedpriortoanalysisinordertoreducethedatasetsize.TheadvantageofusingNRESTsistoavoidoverestimationoftheEST–SSRfrequency.Forexample,afterredundancyelimination, KumpatlaandMukhopadhyay(2005)observed37.3%lossinthenum-berofSSR–ESTfor*Arabidopsisthaliana*.

Forcoffeetrees,wefoundthat6.8%potentialunigenescontainedmicrosatellitemotifs,withanaverageofonemicrosatelliteevery7.73kbofESTsequence.ThisSSR–ESTfrequencywasinthe2.65–10.62%rangeobtainedbyKumpatlaandMukhopadhyay(2005)fordicotspecies.Itwasslightlyhigherthanthe1.5–4.7%rangereportedbyKantetyetal.(2002)formonocots. Theoverallfrequencyandthe frequencyofthediVerentrepeatmotifsareknowntobedependentonthepresence ornotofredundancy,butalsorelatedtothecriteriausedtoidentifySSRinthedatabasemining.Ingeneral,whentheminimumrepeatlengthis20bp,microsatellitesofvariousplantspeciesarepresentinabout5%ofESTs(Varshneyetal.2005a). OurresultsweresomewhatinagreementwiththisestimationalthoughouroverallSSR–ESTfrequencymighthavebeen inXatedbythedetectionofhexanucleodiderepeatmotifsintheanalysis,afactorthatisseldomconsideredinotherstudies(seebelow).

In*C.canephora*ESTs,trinucleotiderepeatswerefoundtobethemostabundant(34.34%),followedbydi-(25.75%)andhexa-nucleotidemotifs(22.04%).

Trinucleotiderepeatsaregenerallythemostcommonmotiffoundinbothmonocots(54–78%amongcereals,Varshneyetal.2005a)anddicots(forexample51.5%inMedicagotruncatula,Eujayletal.2004).Onlyonereport(KumpatlaandMukhopadhyay2005)hasdescribedtheprevalenceofdinucleotiderepeatsinmostofthedicotsinvestigated.However,theysuggestedthattheirresultmighthavebeenduetotheover-representationofuntranslatedregions(UTRs)comparedwithopenreadingframes(ORFs).Indeed,becauseoftheabsenceofframeshiftmutationswhentherearelength variationsintri-andhexanucleotiderepeats,thesemotifsarefoundinexcessinbothcodingandnoncodingsequences,butotherrepeattypesaremuchlessfrequentincodingregionsthaninUTRs(Metzgaretal.2000).Ourresultsconfirmedthisdistribution(Table1):7/9microsatellitesfoundincodingsequencesweretrinucleotiderepeatsandonewasanhexanucleotiderepeat,whereasmicrosatellitesinUTRweremainlydinucleotiderepeats(6/12)andtrinucleotiderepeats(3/12).

Asmostmicrosatellitelibrariesusedformarkerdevelopmentaregenerallyenrichedindi-,tri-andtetra-nucleotiderepeats,computationalminingofESTdatabasesmainlyinvolvesthesetypesofmotifs(Kantetyetal.2002;Morganteetal.2002;Eujayletal.2004;Pintoetal.2004;LaRotaetal.2005).However,anonnegligibleabundanceofmono-andhexanucleotideSSRswasobservedinourstudy,i.e.12.30and22.04%,respectively(Fig.2).Thesedatawereinclose agreementwiththoseofGao(2003),andsupportedtheresultsobtainedbyKumpatlaandMukhopadhyay(2005)andMorganteetal.(2002)ontheabundanceofmononucleotideSSRs. Moreover,whenconsidering*Coffea* microsatelliteswithmono-tohexanucleotiderepeats,theirrelativedistributionmatchedthatnotedinsoybean(Gaoetal.2003;Fig.2).

TheGA/CTmotifwasthemostabundantdinucleotidemotif(62.2%)inour*Coffea* ESTs.Thesemotifswerealso themostfrequentlyobservedSSRs in plants(Scottetal.2000;Gaoetal.2003;Thieletal.2003;Sahaetal.2004).Themostabundanttrinucleotiderepeatmotifdetectedinthepresentstudywas AGG/TCC(23%), closelyfollowed by AAG/TTC (20.3%).Theseresultsareinagreementwithotherreportsondicotspecies(Scottetal.2000;Eujayletal.2004;KumpatlaandMukhopadhyay2005).TherareCCG/GGCfrequencycomparedwithrice(Fig.2)conWrmedthegeneraltrendsnotedinmonocots,i.e.theyhave moreCG-richtrinucleotiderepeatsthandicots(Morganteetal.2002).

OrthologampliWcationandcross-speciestransferability

OurcomputationalstrategytodetectNRSSRsanddevelopuniqueEST–SSRmarkersappearedeYcientsincenearlyallthemarkersappearedtobesinglelocusspeciWc.

Theuse ofa set ofNRsequenceswas important forthedevelopmentofuniquegenicSSRmarkers.ThespeciWcityofthedesignedprimerpairswasthencheckedbyblastingagainsttheESTsequences.However,thepresenceofparalogs(sequencesderivedfromduplica-

tionevents)togetherwithorthologs(sequencesderivedfr

thesourceofdiffculties ininterpreting between-speciescomparisons.The>94%identityofampliconsequencesinSSRflankingregionsandthemaintenanceofrepeatmotifsconfirmedthecross-speciessequenceconservationandtheprimerspecificity.Finally,theorthologywasfurtherconfirmedbyananalysisofamplificationpatterns(sizeandnumb”erofamplificationproducts).

Cross-speciestransferabilitywithin the *Coffea* genuswasveryhighsinceeachprimerpairyieldedaPCRproductinanaverageof6.1species(outofseventested)and14(56%)gaveamplificationwithallthetested*Coffea* species,regardlessofthephylogeneticrelationships.Thisisanimportantfeatureofgenic SSRmarkers,whicharetransferablea mong distantlyrelatedspeciesorevengenera(Decroocqetal.2003;Liewlaksaneeyanawinetal.2004;Varshneyetal.2005b;Sethyetal.2006).Comparedtopreviousresultsbasedongenomicmicrosatellites(Poncetetal.2004),EST–SSRappearedtobemoretransportablemarkers(62.5–92vs.61.7%).Similarly,thetransferabilityofEST–SSRmarkersfrom*Pinustaeda*(loblollypine)to *P.contorta*ssp. *latifolia*wastotal,whileitwasonlylessthanathirdfornon-ESTderivedmicrosatellitemarkers(Liewlaksaneeyanawinetal.2004).

EST–SSRpolymorphism

Ahighlevelofpolymorphicloci wasnotedwhateverthespeciesconsidered.Althoughtheconservednatureof EST–SSRpromotedtransferability,itcouldalsohavelimitedpolymorphism.Thishasbeensuggestedinseveralreportswherethelevelof EST–SSRpolymorphismwaslowerthanthatwithSSRderivedfromgenomiclibraries(Choet al.2000;Eujaylet al.2001;Guptaetal.2003).However,somerecentstudiesreportedhighlevelsofpolymorphismwithEST–SRRmarkers(Eujayletal. 2004;Fraseretal.2004;Sahaet al.2004),withcaseswhereEST–SSRmarkerswereassociatedwithequivalentorevenhigherlevelsofpolymorphismthangenomicSSR(e.g.Liewlaksanee-yanawinetal.2004;Varshneyet al.2005a).Ourresultsonseven*Coffea* speciesalsosupportedtheseobservations.Threeto19allelesperpolymorphiclocuswererecorded amongthesetofspeciestested.Thisrangeislargerthanthatreportedby Bhatetal.(2005)—7–13allelesdetectedusingnineEST–SSR—although awiderspectrumofspecieswastested(14*Coffea* andfour*Psilanthus*).Interestingly,itisalsolargerthanthatnotedbyPoncetetal.(2004),witharound60genomicmicrosatell

omacommonancestor),mightstillbepossibleand itesanalyzedonfour*Coffea* species(2to9alleles).Acomparisonofgeneticparametersestimatedfromthispreviousstudy(Poncetetal.2004)andthe

presentresultscouldbeillustratedwithtwoexamples.First,thecultivatedspecies*C.canephora*displayedhighervalues withEST–SSRthanwithgenomicmicro-satellites, i.e.mean allele numberperlocus (5.7 vs.3.6),observedheterozygosity(0.51vs.0.38)andPIC(0.59vs.0.48).Secondly,comparablevaluesforthesethreeparametersandforthemeanFisvaluewereobtainedfor*C.pseudozanguebariae*independentlyofthetypeofSSRmarker.

Acomparisonofgeneticparametersobtainedforeachspecies illustrated thatEST–SSRmarkerswereidealforassessinggeneticdiversityincoffeetrees.Forexample,*C.pseudozanguebariae*appearedtobelesspolymorphicthan*C.canephora*whenbothgenomic microsatelliteandEST–SSRmarkerswere usedandlesspolymorphicthan*C.dewevrei*and*C.liberica* whenusingEST–SSRmarkers.Thisobservationisinagreementwiththemorerestrictedgeographicaldistributionof*C.pseudozanguebariae*.Whenconsideringthetworelatedspecies*C.liberica*and*C.dewevrei*,theydiVerentiated earlierwhenusingmorphologicaltraits,molecular markers (AFLP)and malefertilityofF_1 hybrids(N'Diayeetal.2005).The*C. dewevrei*individualsanalyzedinthepresentstudywereformerlycultivatedgenotypescollectedintheCentralAfricanRepublic.The*C.liberica*samplecorrespondedto amix ofgeographicoriginscomprisingcultivatedandwildforms. Undertheseconditions,itwasnotsurprisingtoobservehigherlevelsofPICandheterozygosityassociatedwithlowerFisvalues,andanullallelefre-quency(rb).

AnotherimportantfeatureofourEST–SSRs wastheirefficiencyindistinguishingindividualsfrompairs ofrelatedspecies.40to52%diagnosticlociwereidentiWedwhentheWveinterspecificcombinations(CAN/EUG, CAN/HET, CAN/LIB, CAN/PSE and PSE/DEW)wereconsidered.

Finally,ourresultsdemonstratedthat:(1)*C.canephora*EST–SSRmarkerscanbeeasilytransferredtowild*Coffea* speciesforwhich no informationisavailableontheirDNAsequences;(2)theyaregoodcandidatesforthedevelopmentofconservedorthologousmarkersforgeneticanalysisacross*Coffea* species.Thishighdegreeoftransferabilitybetweenspecieswillfacilitatecomparativemappingandhomologousgenecloning.

AcknowledgmentsThisworkwaspartlysupportedbyEUgrantNo.ICA4-CT-2001-10068.TheauthorswishtothankI.Mougenot,C.Fizames,B. Piegu,A.Wissocq,F.Lechauve,F.Moreews, X.Argout,F.Chevalier,andmany GenetropresearchersfortheirinvolvementinthedevelopmentofESTdb,andM.LorieuxforhishelpindevelopingtheSSRscript(http://www.mpl.ird.fr/bio-info/).ThanktoDr.SantiagoC.González-Martínezforhisvalu-ablecommentsonthemanuscript.

References

AltschulSF, Madden TL, SchaVerAA, ZhangJ,ZhangZ,MillerW,LipmanDJ(1997)GappedBLASTandPSI-BLAST:anewgenerationofproteindatabasesearchprograms.Nucle-icAcidsRes25:3389–3402

BaruahA,Naik P, Hendre S,Rajkumar R,RajendrakumarP,AggarwalRK(2003)Isolationandcharacterizationofninemicrosatellitemarkersfrom*Coffea arabica*L.,showingwidecross-speciesampliWcations.MolEcolNotes3:647–650

BhatPR,KrishnakumarV,HendrePS,RajendrakumarP,VarshneyRK,AggarwalRK(2005)IdentiWcationandcharacterizationofexpressedsequencetags-derivedsimplesequencerepeats,markersfromrobustacoVeevariety‘CxR’(aninter-speciWchybridof*Coffea canephora*£*Coffea congensis*).MolEcolNotes5:80–83

BrownGR,KadelEEIII,BassoniDL,KiehneKL,TemesgenB,van BuijtenenJP,SewellMM,MarshallKA,NealeDB(2001)Anchoredreferencelociinloblollypine(*Pinustaeda*L.)forintegratingpinegenomics.Genetics159:799–809

CatoSA,GardnerRC,KentJ,RichardsonTE(2001)ArapidPCR-basedmethodforgeneticallymappingESTs.TheorApplGenet102:296–306

ChoYG,IshiiT,TemnykhS,ChenX,LipovichL,McCouchSR,Park WD,AyresN,CartinhourS(2000)DiversityofmicrosatellitesderivedfromgenomiclibrariesandGenBanksequencesinrice(*Oryzasativa*L.).TheorApplGenet100:713–722

CoulibalyI,RevolB,NoirotM,PoncetV,LorieuxM,Carasco-Lacombe C, Minier J, DufourM, HamonP(2003)AFLPandSSRpolymorphisminA*Coffea* interspeciWcbackcrossprogeny[(*C.heterocalyx*£*C.canephora*)£*C.canephora*].TheorApplGenet107:1148–1155

DecroocqV, FaveMG, HagenL, BordenaveL,DecroocqS(2003)Developmentandtransferabilityofapricotandgrape ESTmicrosatellitemarkersacrosstaxa.TheorApplGenet106:912–922

DirlewangerE,CossonP,TavaudM,AranzanaJ,PoizatC,ZanettoA,ArusP,LaigretF(2002)Developmentofmicrosatellitemarkersinpeach[*Prunuspersica*(L.)Batsch]andtheiruseingenetic diversityanalysisinpeachandsweetcherry(*Prunusavium*L.).TheorApplGenet105:127–138

DufourM,HamonP,NoirotM,RistrerucciAM,BrottierP,VicoV,LeroyT(2001)PotentialuseofSSRmarkersfor*Coffea* spp.geneticmapping.In:ASIC(ed)19thinternationalscienceColloquiumoncoVee,Trieste,Italy

EllegrenH(2004)Microsatellites:simplesequenceswithcomplexevolution.NatRevGenet5:435–445

EujaylI,SledgeMK,WangL,MayGD,ChekhovskiyK,ZwonitzerJC,MianMA(2004)MedicagotruncatulaEST-SSRsrevealcross-speciesgeneticmarkersfor*Medicago*spp.TheorApplGenet108:414–422

EujaylI,SorrellsM,BaumM,WoltersP,Powell W(2001)Assessmentofgenotypicvariationamongcultivateddurumwheatbasedon EST-SSRSandgenomicSSRS.Euphytica119:39–43FernandezD,SantosP,AgostiniC,BonMC,PetitotAS,SilvaMC,GuerraGuimaraesL,RibeiroA,ArgoutX,NicoleM(2004)CoVee(*Coffea arabica*L.)genesearlyexpresseduringinfectionbytherustfungus(*Hemileiavastatrix*).MolPlantPathol5:527–536

FraserLG,Harvey CF,Crowhurst RN,De SilvaHN(2004)EST-derivedmicrosatellitesfromActinidiaspeciesandtheirpotentialformapping.TheorApplGenet108:1010–1016

GaoLF,TangJF,LiHW,JiaJZ(2003)Analysisofmicrosatellitesinmajorcropsassessedbycomputationalandexperimentalapproaches.MolBreed12:245–261

Gonzalez-MartinezSC,Robledo-ArnuncioJJ,ColladaC,DiazA,WilliamsCG,AliaR,CerveraMT(2004)Cross-ampliW-cationandsequencevariationofmicrosatellitelociinEur-asianhardpines.TheorApplGenet109:103–111
GonzaloMJ,OliverM,GarciaMasJ,MonfortA,DolcetSanjuanR,KatzirN,ArusP,MonforteA(2005)Simple-sequencere-peatmarkersusedinmerginglinkagemapsofmelon(*Cucumismelo*L.).TheorApplGenet110:802–811
GuptaPK,RustgiS,SharmaS,SinghR,KumarN,BalyanHS(2003)TransferableEST-SSRmarkersforthestudyofpoly-morphismandgeneticdiversityinbreadwheat.MolGenetGenomics270:315–323
KanteryRV,LaRotaM,MatthewsDE,SorrellsME(2002)Datamaningforsimplesequencerepeatsinexpressedsequencetagsfrombarley,maize,rice,sorghumandwheat.PlantMolBiol48:501–510
KumpatlaSP,MukhopadhyayS(2005)Miningandsurveyofsim-plesequencerepeatsinexpressedsequencetagsofdicotyle-donous species.Genome48:985–998
KyCL,BarreP,LorieuxM,TrouslotP,AkaVouS,LouarnJ,CharrierA,Hamon S,Noirot M(2000)InterspeciWcgeneticlinkagemap,segregationdistortionandgeneticconversionincoVee(*Coffea* sp.).TheorApplGenet101:669–676
LaRotaM,KanteryRV,YuJK,SorrellsME(2005)Nonrandomdistributionandfrequenciesofgenomicand EST-derivedmicrosatellitemarkersinrice,wheat,andbarley.BMCGe-nomics6:23
LashermesP,CombesMC,PrakashNS,TrouslotP,LorieuxM,CharrierA(2001)Geneticlinkagemapof*Coffea canephora*:eVectofsegregationdistortionandanalysisofrecombinationrateinmaleandfemalemeioses.Genome44:589–596
ParkYH,AlabadyMS,UlloaM,SicklerB,WilkinsTA,YuJ,StellyDM,KohelRJ,el-ShihyOM,CantrellRG(2005)Ge-neticmappingofnewcottonWberlociusingEST-derivedmi-crosatellitesinaninterspeciWcrecombinantinbredlinecottonpopulation.MolGenetGenomics274:428–441
PintoLR,Oliveira KM,UlianEC,GarciaAA,deSouzaAP(2004)Surveyinthesugarcaneexpressedsequencetagdata-base(SUCEST)forsimplesequencerepeats.Genome47:795–804
PoncetV,HamonP,MinierJ,Carasco-LacombeC,HamonS,NoirotM(2004)SSRcross-ampliWcationandvariationwith-incoVeetrees(*Coffea* spp.).Genome47:1071–1081
RalloP,TenzerI,GesslerC,BaldoniL,DoradoG,MartinA(2003)Transferabilityofolivemicrosatellitelociacrossthegenus Olea. TheorApplGenet107:940–946
RovelliP,MettulioR,AnthonyF,AnzuetoF,LashermesP,GraziosiG(2000)Microsatellitesin*Coffea arabica*L.In:SeraT,SoccolCR,PandeyA,RoussosS(eds)CoVeebio-technologyandquality. Kluwer,Netherlands, pp123–133
RozenS,SkaletskyHJ(2000)Primer3ontheWWWforgeneralusersandforbiologist programmers. In:KrawetzS,MisenerS(eds)Bioinformaticsmethodsandprotocols:methodsinmolecularbiology.HumanaPress,Totowa, pp365–386
SahaMC,MianMA,EujaylI,ZwonitzerJC,WangL,MayGD(2004)Tallfescue EST-SSR markers with transferabilityacrossseveralgrassspecies.TheorApplGenet109:783–791
SahaMC,MianR,ZwonitzerJC,ChekhovskiyK,HopkinsAA(2005)AnSSR-andAFLP-basedgeneticlinkagemapoftallfescue(*Festucaarundinacea*Schreb.).TheorApplGenet110:323–336
LashermesP,CombesMC,TrouslotP,CharrierA(1997)Phylogeneticrelationshipsofcoffee-treespecies(*Coffea* L.)asinferredfromITSsequencesofnuclearribosomalDNA.TheorApplGenet94:947–955
LemP,LallemandJ(2003)GrassconsensusSTSmarkers:aneY-cientapproachfordetectingpolymorphismin*Lolium*.TheorApplGenet107:1113–1122
LiewlaksaneeyanawinC,RitlandCE,El-KassabyYA,RitlandK(2004)Single-copy,species-transferablemicrosatellitemark-ersdevelopedfromloblollypineESTs.TheorApplGenet109:361–369
LinC,MuellerLA,CarthyJM,CrouzillatD,PetiardV,TanksleySD(2005)CoVeeandtomatosharecommongenerepertoiresasrevealedbydeepsequencingofseedandcherrytranscripts.TheorApplGenet112:114–130
LiuK,MuseSV(2005)PowerMarker:anintegratedanalysisenvironmentforgeneticmarkeranalysis.Bioinformatics21:2128–2129
MetzgarD,BytofJ,WillsC(2000)Selectionagainstframeshiftmutationslimitsmicrosatelliteexpansionincoding DNA.GenomeRes10:72–80
MoncadaP,McCouchS(2004)Simplesequencerepeatdiversityindiploidandtetraploid*Coffea* species.Genome47:501–509
MorganteM,HanafeyM,PowellW(2002)Microsatellitesarepreferentiallyassociated withnonrepetitive DNAinplantgenomes.NatGenet30:194–200
N'DiayeA,NoirotM,HamonS,PoncetV(2006)Geneticbasisofspeciesdi Verentiationbetween*Coffea libérica*Hiernand*C.canephora*Pierre:analysisofaninterspeciWccross.GenetResourCropEvol(inpress)
N'DiayeA,PoncetV,LouarnJ,HamonS,NoirotM(2005)Ge-neticdi Verentiationbetween*Coffea libérica*var.*libérica*and*C.libérica*var*Dewevrei*andcomparisonwith*C.canephora*.PlantSystEvol253:95–104

SambrookJ,FristchEF,ManiatisT(1989)Molecularcloning–alaboratorymanual.ColdSpringHarborLaboratoryedn.ColdSpringHarbor

ScottKD,EgglerP,SeatonG,RossettoM,AblettEM,LeeLS,HenryRJ (2000) Analysis ofSSRsderivedfromgrapeESTs.TheorApplGenet100:723–726

SethyNK,ChoudharyS,ShokeenB,BhatiaS(2006)IdentiWcationofmicrosatellitemarkersfromCicerreticulatum:molecularvariationandphylogeneticanalysis.TheorApplGenet112:347–357

ThielT,MichalekW,VarshneyRK,GranerA(2003)ExploitingESTdatabasesforthedevelopmentandcharacterizationofgene-derivedSSR-markersinbarley(*Hordeumvulgare*L.).TheorApplGenet106:411–422

VarshneyRK,GranerA,SorrellsME(2005a)Genicmicrosatel-litemarkersinplants:featuresandapplications.TrendsBiotechnol23:48–55

VarshneyRK,SigmundR,BornerA,KorzunV,SteinN,SorrellsME,LangridgeP,GranerA(2005b)InterspeciWctransferabilityandcomparativemappingofbarleyEST-SSRmarkersinwheat,ryeandrice.PlantSci168:195–202

VigourouxY,MitchellS, MatsuokaY, HamblinM,KresovichS,SmithJS,JaquethJ,SmithOS,Doebley J(2005)Ananalysisofgeneticdiversityacrossthemaizegenomeusingmicrosat-ellites.Genetics169:1617–1630

WuKS,TanksleySD(1993)Abundance,polymorphismandgeneticmappingofmicrosatellitesinrice.MolGenGenet241:225–235

YuJK,LaRotaM,KantetyRV,SorrellsME(2004)ESTderivedSSRmarkersforcomparativemappinginwheatandrice.MolGenet Genomics271:742–751