



thuk&okbM , l k&l , 'ku LVMht ¼h MGY; w, , l ½eab; W\$ku rduhd&edse&qs , oap&kr; k&, d l eh&lk

jkgg cut&h] H&jrh] 'k&kuk c&e&] i&dt nkl 'j r&k&hj vgen'

l kj&k&k

thuk&okbM , l k&l , 'ku LVMh ½GWAS½ e&ach&ejh l st&M& t&u&Vd dh [k&st gr&q0; f&ä; k&dsMh, u, ek&d& l Z&dk&L&d&u fd; k t&rk g& tc u, t&u&Vd l &kk&dh ig&ku g&rh g& r&sb&l tkud&jh dk mi; k&x ch&ejh dh ig&ku] by&kt v&g& j&k&d&F&k&e gr&q&gr&j j. kuh&r; kacukuse&enn feyrh g& b&E&l; W\$ku t&u&Vd v/; ; uk&e&av&u&V&k&b&IM t&u&ks&/k&bi dk i&mk&Z&ek&u fd; k t&rk g&st c& M&S&/k x&q&ko&U&kk] y&xr] f&M&t&k&bu l e&L; kv&ka&ds&dk&j. k vu&i&y&C&/k g&rk g& ; g , d &ek&f. kr l k&ä; dh rduhd g&st&l dk mi; k&x ?kuh t&u&ks&/k&bi t&kp i&S&y l sg&ys&ks&/k&bi l x&e&W m/k&j. k dj&ds&vun&[k&st&u&ks&/k&bi dk vk&dyu dj&us&gr&q&f&d; k t&rk g& t&u&ks&/k&bi b&E&l; W\$ku thuk&e&0; ki d l &kv&u L&d&u ds&fo'y&S&. k dj&use&eg&roi w&k&Z&g&ä bl 'k&ks& i = e&l &ki e&af&e&l & M&S&/k l e&L; kv&ka&v&g& f&of&h&ku b&E&l; W\$ku f&of&/k; k&ä&ks&n' k&Z& k x; k g&ä

"&on d&f&h&ch&y] Q&k&L&Q&ä] thuk&e& okbM , l k&l , 'ku LVMht] b&ä; W\$ku ds&r&j&h& b&E&l; W] e&ß] f&e&l & e&f&su&Te&A

Issues and Challenges of Imputation Techniques in Genome Wide Association Studies (GWAS): A Review

Rahul Banerjee¹, Bharti¹, Shbana Begum², Pankaj Das¹, Tauqueer Ahmad¹

10.18805/BKAP597

ABSTRACT

A genome-wide association study (GWAS) rapidly scans DNA markers in many individuals to find genetic links to diseases. New findings aid in disease detection, treatment and prevention. Imputation predicts untyped genotypes in genetic studies when data is missing due to quality, cost, or design issues. It's a proven statistical technique for estimating unobserved genotypes by borrowing haplotype segments from a densely genotyped reference panel. This allows estimation and testing of associations at unassayed variants. Genotype imputation is vital in analyzing genome-wide association scans, helping geneticists evaluate evidence for association at untyped genetic markers. This summary outlines missing data issues and various imputation methods.

Key words: BEAGLE, fastPHASE, Genome wide association studies, Imputation methods, IMPUTE, MACH, Missing mechanisms.

vk&d&M&e&ß] f&e&l & M&S&/k rc g&rk g&st c& f&d&l h voy&k&du e&ä pj ds&fy, d&k&b&Z& M&S&/k e&ku l &g&hr u&gh& fd; k t&rk g&ä f&e&l & M&S&/k , d l k&e&U; ?&v&u&k g&S&v&g& vu&u& ä&k&ku dh 'k&ß&v&kr ds&cn l s' k&ks&dr&ä&/k&ä&ks&p&u&k&h nh g&ß fo'k&S& : i l svu&q&ä&Z& vu&u& ä&k&ku ds&fy,] ft&l e&ä , d gh 0; f&ä ij c&gr l k&js vk&d&M&e& g&rs g&ä 'k&ks&dr&ä&/k&ä }&k&j&k mi; k&x dh tk&us& ok&yh &f&Ø; kv&ka&dk&se&ß; : i l sch&l o&h&'kr&k&C&nh e&af&od&f&l r fd; k x; k F&k& ft&l s&l ä&w&Z& M&S&/k ds&fy, f&od&f&l r fd; k x; k g&ä f&e&l & M&S&/k l k&ä; dh; vu&ek&u&ka dh o&S&rk ij &f&rd&y &h&ko M&y l dr&sg&ä

¹ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, Pusa-110 012, New Delhi, India.

²ICAR-National Institute for Plant Biotechnology, LBS Centre, Pusa Campus-110 012, New Delhi, India.

Corresponding Author: Rahul Banerjee, ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, Pusa-110 012, New Delhi, India. Email: rahuliasri@gmail.com

How to cite this article: Banerjee, R., Bharti, Begum, S., Das, P. and Ahmad, T. (2023). Issues and Challenges of Imputation Techniques in Genome Wide Association Studies (GWAS): A Review. Bhartiya Krishi Anusandhan Patrika. DOI: 10.18805/BKAP597.

Submitted: 21-09-2022 **Accepted:** 23-09-2023 **Online:** 13-10-2023

thuke okBM , l kfi , 'ku LVMht ½GWA½

vkupfi'kdh eþ thuke&okBM , l kfi , 'ku LVMh ½GWA½ LVMh GWAS½ ftls l a w k thuke , l kfi , 'ku LVMh ½GWA LVMh WGAS½ ds: i eaHkh tkuk tkrk gþ tksd fofHku 0; fä; kaeavkupfi'kd ofj, ½ ds thuke&okBM l ½ dk , d ijh{k.k gþ; g nskusdsfy, fd; k tkrk gþ fd D; k dkbz çdkj gþ, d fo'kkrk dsl kfk tþMk gþk gþ GWAS vkerkj ij , dy&U; fDy; k/kbM cgg irkvka¼, l , uih½ vks çedk ekuo jkskadsy{k.kkadschp l ½kksij /; ku dæer djrs gþ yfdu l eku : i l sfdl h Hkh vU; tho ij ykxwfd; k tk l drk gþ thuke&okBM , l kfi , 'ku LVMht oKkfudka ds fy, ekuo jks ea 'kkfey thu dh igpku djusdk , d vi{k-k-r u; k rjhdg gþ ; g fof/k thuke dksNks/scnykoka dks [kkstrh gþ ftlgaf axy U; fDy; k/kbM i , yhe, fQZe ; k , l , uih dgk tkrk gþ tksfcuk chekjh okys ykska dh rgyuk eafdl h fo'kks chekjh okys ykskaeavf/kd ckj gkrs gþ çR; d v/; ; u , d gh l e; ea l sMka ; k g tkjka , l , uih nsk l drk gþ 'kkkdrkzbl çdkj dsv/; ; u ds M/k dk mi; kx mu thukadksbæx djusdsfy, djrs gþ tks fdl h 0; fä ds , d fuf'pr chekjh ds fodkl ea ; kx nku dj l drsgþ tc ekuo M/k ij ykxwfd; k tkrk gþ rksGWA LVMht çfrHkfx; kadsMh, u, dh rgyuk fdl h fo'kks y{k.k ; k chekjh dsfy, vyx&vyx Qusk/kbi okys djrs gþ ; s çfrHkxh , d chekjh okys yks , oa fcuk ½fu; æ.k½ ds l eku yks gks l drs gþ ; k os fdl h fo'kks y{k.kkadsfy, vyx&vyx Qusk/kbi okys yks gks l drs gþ mnkgj.k ds fy, jækpi A bl -f"Vdks k dks Qusk/kbi &çFke ds: i ea tkuk tkrk gþ ftl eaçfrHkfx; ka dks igys muds uñkfud vfHk0; fä; ka }kjk oxh-z-r fd; k tkrk gþ ts k fd thuk/kbi &çFke ds foijhr gkrs gþ çR; d 0; fä Mh, u, dk , d uenuk nrk gþ ftl eal syk [kka vkupfi'kd : i ka dks , l , uih l kjf.k; ka dk mi; kx djd i < k tkrk gþ ; fn chekjh okys ykskaea, d , yhy vf/kd i k; k tkrk gþ rksbl , yhy dks chekjh l s tþMk gþk dgk tkrk gþ dkbz Hkh nks ekuo thuke yk [kka vyx&vyx rjhdka l s fHku gkrs gþ thuke ¼ l , uih½ ds vyx&vyx U; fDy; k/kbM eaNks/scnyko gkrs gþ vks l kfk gh dbz cMscnyko gkrs gþ ts sfoyki u l fëy vks çfrfyi l æ; k ½d, i h uej½ fHkuurk, A o"z 2000 ds vkl ikl j thMCY; w v/; ; uka dh 'kævkr l sigyþ tkp dk çkFkfed rjhdg

ifjokjkaevupfi'kd l ædkadsoakkuþe v/; ; u dsek/; e l sfd; k tkrk FkA ; g -f"Vdksk , dy thu fodkjka ds fy, vR; f/kd mi; kxh l kfr gþk FkA gkykfd] tfVy chekfj; kadsfy, vkupfi'kd fyðst v/; ; u dsifj. kkeakds i þ% i sk djuk ef' dy l kfr gkrs gþ fyðst v/; ; u dk , d l ç-k; k x; k fodYi vkupfi'kd , l kfi , 'ku LVMh gþ ; g v/; ; u tkprk gþ fd D; k os 0; fä; k ftuds ikl , yhy #fp l s l æd/kr Hkkrd xqk gkrs gþ mlgavkupfi'kd : i karj dsl mHkZ eavifçkr l svf/kd ckj nsk tkrk gþ

i) fr

th MCY; w, v/; ; ukadk l cl svke -f"Vdksk d d &d/ky l ½vi gþ tks0; fä; kadsnks cMsc l engkþ , d LoLFk fu; æ.k l eng vks , d chekjh l s çHkfor] d l eng dh rgyuk djrk gþ , l , uih dh l Vhd l æ; k thuk/kbfi æ rduhd ij fuHkj djrh gþ yfdu vkerkj ij ; g , d fefy; u ; k vf/kd Hkh gks l drh gþ bueal çR; d , l , uih dh tkp dh tkrh gþ vks nsk tkrk gþ fd D; k , yhy vkofuk mipkj , oa fu; æ.k l eng ds chp egRo i w k : i l s cny tkrh gþ v, M+ vuq kr] nks v, M+ dk vuq kr gþ tks GWA v/; ; ukadsl mHkZ ea, d fof'kV , yhy okys0; fä; ka dsfy, jks dh l Hkkouk gsvks mu 0; fä; kadsfy, jks dh l Hkkouk gsfuds ikl l eku , yhy ugha gþ

xf.krh; : i l s

ekuk A dkbz ?kVuk gS rc] fdl h ?kVuk ds ?kVr gkrs dh l Hkkouk ¼ P½ ?kVuk ?kVrh gS P½ ?kVuk ?kVr ugha gkrs h ¼

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1-P(A^c)}$$

, d mnkgj.k yrs gþ

	, yhy dkm/	
	G	T
ds d	a	b
fu; æ.k	c	d

uenusea l Hkh g , yhy ij fopkj djþ vks ; k-fPNd : i l s, d pþA , d ekeysea g , yhy gkrs dh l Hkkouk %^a A uenusea l Hkh t , yhy ij fopkj djþ vks ; k-fPNd : i l s, d pþA , d ekeysea , d Vh t , yhy gkrs dh l Hkkouk %^b A

v,Mt vuqkr 1/2OR1/23/4 ad/bc

0; k[; k bl çdkj dgh tk l drh g%

OR3/4 çR; d vfrfjä g ,yhy dsfy, dš gksdsh l Hkkouk ea of) A

OR3/41%thukv/bi vks jks dschp dkbz l ædk ughA

OR>1%G ,yhy l schekjh dk [krjk c<+tkrk gA

OR<1%T ,yhy l schekjh dk [krjk c<+tkrk gA

tc dš xij ea ,yhy vkofük fu; æ.k xij dh rgyuk eacgr vf/kd gkrh gš rksv,Mt vuqkr 1 l svf/kd gkrk gš vks de ,yhy vkofük ds fy, bl ds foijhrA l ds vfrfjä] v,Mt vuqkr ds egro ds fy, P&eku dh x.kuk vke rks ij , d l k/kkj.k dkb&LDoMZ1/42/2 V&V dk mi; kx djdsdh tkrh gA thMCY; w v/; ; u dk mıs; 1 l s dkQh fhkuu v,Mt vuqkrka dk irk yxkuk gSD; kfd bl l sirk pyrk gsfed , l ,uih chekjh l stfMk gS; k ughA thukv/bi bā; w'sku l k[; dh; fof/k; ka }kjk fd; k tkrk gS tksGWAS Mv/k dks gsyk/bi dsl nHkziSy dsl kfk tkMfS gA thukv/bi bā; w'sku dsfy, ekstmk l ,fVos j i dslst ea IMPUTE2 vks Mach 'kkfey gA

febl æ Mv/k dsl k

l ožk.k vuq dku ea febl æ Mv/k ds rhu e[; l k g% u,u&dojst] dš xš&çfrfØ; k 1/2u,u j i k d 1/2 vks vkbVe xš&çfrfØ; k 1/2kbl , V vy] 20041/2 u,u&dojst rc gkrk gS tc dN l efV ds uews ea paps tkus dh dkbz l Hkkouk ughagkrh gA febl æ Mv/k xš&çfrfØ; k l svkrk gS tc dkbz çfroknh l ožk.k ij fdl h Hkh vkbVe dk tokc nus l s budkj djrk gA vkbVe xš&çfrfØ; k rc gkrh gS tc l ožk.k ij dšy vki'kd vkbVe ijk fd; k tkrk gA yki rk Mv/k dsbu rhu l kkaep xš&dojst vks dšy xš&çfrfØ; k dksmi; æ uewuk Hkj dk mi; kx djds l ækS/kr fd; k tk l drk gS tks uewuk dks yf{kr l efV dk l Vhd : i l s çfruf/kRo djus ds fy, fMtkbu fd; k x; k gS 1/2phk] 20141/2 gkykfd] vkbVe xš&çfrfØ; k ds dkj.k febl æ eku otu dk mi; kx djdsr; ughafd, tk l drsgA vkbVe xš&çfrfØ; k okysekeykadsl phokj gVkus l sdN eV; oku tkudkjh dk upl ku gkrk gS vks l Hkkfor : i l s ck; l i šk gkrh gA febl æ Mv/k dsl kfk , d e[; fpar; g gsfed D; k i wZ Mv/k oky uewuk vHkh Hkh yf{kr vkcknh l efV dk çfruf/k gS 1/2j kfk] 19941/2

febl æ ræ

febl æ Mv/k ræ dks vke rks ij rhu l egka ea oxhZ-r fd; k tkrk gS 1/2fcu] 19761/2 ; k-fPNd : i l siyh rjg l s xk; c 1/2MCAR1/2 ; k-fPNd : i l s xk; c 1/2MAR1/2 , oa ; k-fPNd : i l s xk; c ugha 1/2MNAR1/2 ; fn ftu fo"ka; ka ea vkcdMs xep gš os fo"ka; ka ds ijs uews dk , d ; k-fPNd mi l epp; gš rks febl æ Mv/k dks iyh rjg l s ; k-fPNd dgk tkrk gS 1/2MCAR1/2 dšy i wZ Mv/k dk mi; kx djus l fu"i {k ifj.kke fey l drk gA MAR rc gkrk gS tc xep gksk çf{kr Mv/k ij fuHkj ugha djrk gA Mv/k MNAR rc gkrsga tc fdl h pj ij Mv/k xep gksdsh l Hkkouk ml pj dseku ij fuHkj djrh gA

vkuof'kd egkekjh foKku eafebl æ Mv/k

vkuof'kd egkekjh foKku ds v/; ; u ep Mv/k dh xqkoÜkk] ykx n{krk ; k rduhdh fMtkbu ds dkj.k kads fo'yšk.k ds fy, fo'kšk ekdj ka ds thukv/bi vuqyC/k gks ij febl æ Mv/k l eL; k, amRi lu gkrh gA l cl se[; thukv/bi bā; w'sku fof/k; ka ea IMPUTE, FastPHASE, Mach vks BEAGLE 'kkfey gš vekfpZuh vks gkoh] 20101/2 ekuo vkuof'kd egkekjh foKku dk mıs; fof'kV Qauk/bi l s l æ/kr jks ; k jks&l ækh y{k.k ij vkuof'kd : i karjka dh igpku djuk gA thukv/bi æ v/; ; u geškk p; fur ekdj l vka rd gh l hfer jgk gš e[; : i l s , dy U; fDy; k/kbM cgq irk 1/4 l ,uih 1/2 tks0; kol kf; d : i l s mi yC/k ekbØks jsij Lo: fir gkrsgA fyndst vl kE; oLFkk ij vk/kkfjr thukæ&okbM , l kfi , 'ku LVMht 1/2thMCY; w , l 1/2 dšy , l ,uih dh , d Nks/h l [; k dk mi; kx djrk gS tks fdl h nh xbZ l efV ea 80% vkuof'kd fhkuurk dk fooj.k djrh gS 1/2Ø, Ld] 20151/2 5 thukv/bi bā; w'sku fof/k; k oSkfudka dks febl æ Mv/k l eL; k dks l ækS/kr djusea enn dj l drh gš 1/2ekfpZuh , M gkoh] 20101/2

bEl; w'sku fof/k; k

egkekjh foKku ds v/; ; u ds fy, fofHkuu thukv/bi bā; w'sku fof/k; ka dk çLrko fd; k x; k gA mu thukv/bi bā; w'sku fof/k; kadschp e[; varj ; g gsfed , d , l ,uih dsl 'krZ thukv/bi forj.k dks dš si fjhkkf'kr vks mi; kx fd; k tkrk gA eku ya fd gekjs ikl L Mk; yfyd v,Vkd key , l ,uih ea Mv/k gS vks çR; d , l ,uih eank

, yhy dks 0 vks 1 l s dksMr fd; k x; k gA H] bu L
 , l , uih eaN gsyk/kbi ds, d l v dksn'kkzrk gStcfd
 g] L , l , uih ea thuk/kbi Msk dk l v gA thuk/kbi
 bh; Wsku dk mrs; mu , l , uih dsthuk/kbi dh Hkfo"; ok.kh
 djuk gSftlgav/; ; u uenuea thuk/kbi ughafd; k x; k
 gA ; gk mfYyf[kr l Hkh bh; Wsku fof/k; ka ekurh gS fd
 febl x Msk MAR gS, oal kjsHMM vk/kkfjr bh; Wsku fof/k; k
 gS ; g l nHkz iSy ij l 'krZ v/; ; u eafy, x, cR; d
 0; fa ds febl x thuk/kbi dk vuoku yxkrs gA , d k
 djus l nHkz iSy dk mi; kx cR; d 0; fa dseVhykd l
 thuk/kbi eapj.k vfuf'prrk ij fo'kyk.kkRed : i l s
 , dh—r djus ds fy, fd; k tkrk gA

bel; W IMPUTE½

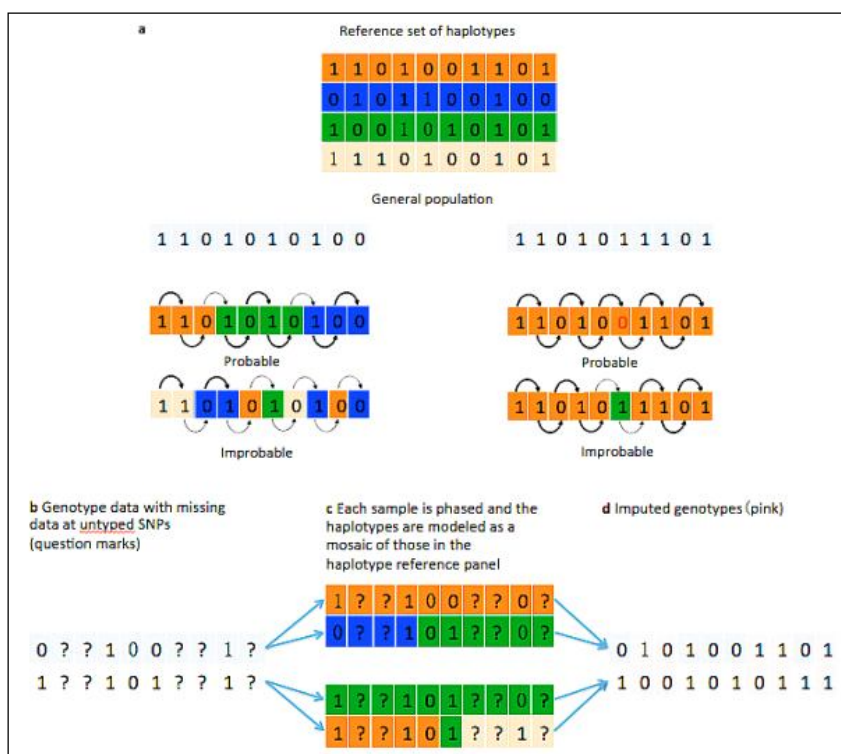
ekfpzh , V vy ½2007½ usIMPUTE fof/k dk igyk l dj.k
 rS kj fd; k ½fp= 1½A IMPUTEV1 l gl a ksth o[kadsvu d j.k
 ds fy, fgMau ekdkb e, My ½HMM½ ds foLrkj ij vk/kkfjr
 gA ½Qs jugM , oaMksy h] 2001½ vks fydst fmfl fi fyfcz e
 ds e, Mfyax vks i q l i kst u njka ds vkdyu ds fy,

½y h , M LVhQd] 2003½A ; g fof/k cR; d 0; fa dsthuk/kbi
 oDVj G, ds, p, e, e ij vk/kkfjr gStksN Kkr gsyk/kbi
 H ds, d l v ij l 'krZ gA , d fNis gq ekdkb e, My
 ¼ p, e, e½ dk : i bl cdkj gkrk g%

$$\Pr(G_i | H) = \sum_{Z_i^{(1)} Z_i^{(2)}} \Pr(G_i | Z_i^{(1)}, Z_i^{(2)}, H) \Pr(Z_i^{(1)}, Z_i^{(2)}, H)$$

tgk Z_i⁽¹⁾ = {Z_{i1}⁽¹⁾, ..., Z_{iL}⁽¹⁾} vks Z_i⁽²⁾ = {Z_{i1}⁽²⁾, ..., Z_{iL}⁽²⁾} L
 l kbVka ij fgMau LVs ds nks Øe gS vks Z_i⁽⁰⁾ ∈ {1, ..., N} Z_i
 dks l nHkz iSy H l gsyk/kbi dh tkMh ds : i eaekuk
 tk l drk gSftls thuk/kbi oDVj G, cukus ds fy,
 d, ih fd; k tkrk gA Pr(G_i⁽¹⁾, Z_i⁽²⁾, H); g i mZ l Hkkouk dks
 ifjHkkf"kr djrk gSfd ds sd, ih fd, x, gsyk/kbi l dh
 tkMh vuØe ds l kFk cnryh gS vks , d ekdkb Jdkyk
 Jkj ifjHkkf"kr dh tkrh gSftl eafLo fpa nj thuke Hkj
 ea Bh d&i ekus ij i q l i kst u ekufp= ds vuoku ij
 fuHkZ djrh gA ekdkb Jdkyk dh ckjHkd voLFkk N²
 LVs ij , d l eku gA

$$\Pr(G_i^{(1)}, Z_i^{(2)}, H) = \frac{1}{N^2}$$



fp= 1%Øotkd ½2015½ ij vk/kkfjr bh; W ds l kFk thuk/kbi bh; WskuA

ge l kbV / l s/+1 rd J[kyk dh l Øe.k l Hkkoukvka dh x.kuk dj l drsg%

$$\Pr(\{Z_{il}^{(1)}, Z_{il}^{(2)}\} \rightarrow \{Z_{i(l+1)}^{(1)}, Z_{i(l+1)}^{(2)}\}, H) = \begin{cases} (e^{-\frac{\rho_l}{N}} + \frac{1 - e^{-\frac{\rho_l}{N}}}{N})^2, Z_{il}^{(1)} = Z_{i(l+1)}^{(1)}, Z_{il}^{(2)} = Z_{i(l+1)}^{(2)} \\ (e^{-\frac{\rho_l}{N}} + \frac{1 - e^{-\frac{\rho_l}{N}}}{N}) (\frac{1 - e^{-\frac{\rho_l}{N}}}{N}), Z_{il}^{(1)} = Z_{i(l+1)}^{(1)}, Z_{il}^{(2)} \neq Z_{i(l+1)}^{(2)}, Z_{il}^{(1)} \neq Z_{i(l+1)}^{(1)}, Z_{il}^{(2)} = Z_{i(l+1)}^{(2)} \\ (e^{-\frac{\rho_l}{N}} + \frac{1 - e^{-\frac{\rho_l}{N}}}{N}), Z_{il}^{(1)} \neq Z_{i(l+1)}^{(1)}, Z_{il}^{(2)} \neq Z_{i(l+1)}^{(2)} \end{cases}$$

$\rho_l = 4N_e r_l$ v[r_l l kbVka/ v[r_l /+1 dschp çfr ih-k vkupf'kd njih gA imzforj.k ds: i eafy[kk tk l drk g%

$$\Pr(Z_i^{(1)}, Z_i^{(2)}, H) = \Pr(Z_{il}^{(1)}, Z_{il}^{(2)}, H) \prod_{i=1}^{L-1} \Pr(\{Z_{il}^{(1)}, Z_{il}^{(2)}\} \rightarrow \{Z_{i(l+1)}^{(1)}, Z_{i(l+1)}^{(2)}\}, H)$$

; g] $\Pr(G_i | Z_i^{(1)}, Z_i^{(2)}, H)$ i fjHkkf"kr djrk gSfd n[ksx, thukv/bi d[sdjhc gkx[y[du Bhd ml h rjg ughat[s gsyk/bi dh udy dh tkrh gA

bl fp= ea0 v[1 , d l mHkZ, yhy dh mi fLFkr ; k vuq fLFkr dks n'kkZrk g[bh; W[s ku d[%gsyk/bi dk l mHkZ l %pj.kc) gSv[bl eapkj vyx&vyx gsyk/bi 'kkfey g[$1/2$ tul [; k gsyk/bi dks l mHkZ l % l gsyk/bi dk ekst[ekuk tkrk gA l Øe.k l Hkkoukvka ds l kFk ekdk[J[kyk e, My }kjk i fjHkkf"kr l m[f kr gsyk/bi forj.k tul [; k bfrgkl v[LF kuh; i q[l kstu ekufp= nksukaij fulH[djrk gA ckM rhj mPp l Øe.k l Hkkoukvka dks b[x r djrs g[tcf[yxrk gSfd rhj de l Øe.k l Hkkoukvkadk çfrfuf/kRo djrs gA yky dk vFkZgSE; W[s kuA %ch[f cuK Vbki fd, x, , l , uih %ç'u fp[$1/2$ ij ykirk M[s /k ds l kFk thukv/bi M[s /k dks n'kkZrk gA % h[$1/2$ çR; d[uemusd[spj.kc) fd; k tkrk gSv[$gsyk/bi$ dks gsyk/bi l mHkZ i %y eaekst[ds: i ear[kj fd; k tkrk gA %h[$1/2$ v/; ; u ds uemus ea fefl æ thukv/bi dks l mHkZ l % ea feyku djus okys gsyk/bi dk mi ; k[x djds yxk; k tkrk gA l k[W ; r[g ij] fefl æ M[s /k ds l kFk , d pj.kc) thukv/bi dh l Hkkouk dk e[W ; k[W u ekst[gsyk/bi ds l Hkh l Hkkfor tkM[s ij fopkj djdsfd; k tkrk gS tksn[ks x, M[s /k ds l kFk l ær gk[W v[g l clsl Hkkfor tkM[s x[g & Vbki fd, x, , l , uih ij l clsl Hkkfor thukv/bi fu/kkZjr djrh gA IMPUTEv2 dks yh , oaLVhQd %2003%

}kjk fodfl r fd; k x; k g[s tks IMPUTEv1 dh rgyuk ea T; knk vPNk -f"Vdksk gA çR; d[0; fä dsfy, , d vyx] fo'y[s .kkRed bEl; W[$pj.k$ djus ds ctk;] IMPUTEv2 l Hkh 0; fä; k[W ds, d i q[$koUk$ <kp[sea , d l kFk yxkrk gA , l , uih dks v/; ; u ds uemus v[g l mHkZ i %y % $1/2$ Vh[2 nksukaea thukv/bi fd; k tkrk gS l QZ l mHkZ i %y % $1/2$ V[2 ; % ea thukv/bi fd; k tkrk g[s bl vk/kkj ij , l , uih dksnks vyx&vyx l %kaefoHkkftr fd; k tkrk gA v/; ; u ds uemus % $1/2$ V[2 Vh[2 ea, l , uih ij gsyk/bi dk vu[ku igys yxk; k tkrk g[s v[g fQj ; w[ea , l , uih ea , yhy dks or[ku ea vu[ku fur gsyk/bi ij l 'krZyxk; k tkrk gA fo'k[s : i l } ; g , d ekdk[J[kyk ek[s dky[$1/2$ el h, el h[2 , Yx[g jne pykrk gS tksnksçfu; knh pj.kk[W dschp o[df yi d gk[rk g[$1/2$ pj.k l Hkh n[ks x, thukv/bi v[g v/; ; u uemus ea, l , uih eafdl h Hkh fNVi[W : i l sykirk thukv/bi dks yxw[dja % $1/2$ V[2 Vh ea, l , uih % fdl h fn, x, , l , uih ea thukv/bi fd, x, l Hkh 0; fä; k[W dsfy, i[y pj.k dh tkudkj; 2% fi Nys pj.k ea vu[ku fur çR; d[gsyk/bi dsfy,] x[g & Vbki fd, x, , l , uih %gkoh v[g ekfp[uh] 2009% eaykirk , yhy l dks yxw[$djus$ dsfy, l mHkZ i %y dk mi ; k[x d[W gsyk/bi l dsfy, , d l Hkk0; rk forj.k dks IMPUTEv1 ds l eku i fjHkkf"kr fd; k x; k g[s v[g nks gsyk/bi l eamPpre i'p çk[f ; drk gS tks v/; ; u uemus dsfy, p[us x, gA IMPUTEv2 IMPUTEv1 dh rgyuk ea cg[r rst gS D; k[W d bh; W[s ku pj.k vx[q .kr çfr: i.k g[s

IMPUTEv1 eao 1/n2½ l s IMPUTEv2 eao 1/n2 rd x.kuk
l e; dksde djrk gS¼n l kkkfor gS½k/kbi dh l ½; k dks
n'kk½k gS ½gkoh½ Mkus½h v½½ ekfp½h½ 2009; ekfp½h½ v½½
gkoh½ 2010½A

QKLVQSt 1/4last PHASE1/2

;g fof/k , l , uih gšyK/kbll ij vk/kkfjr gš tks l eku
gšyK/kbll ¼ khV , M LVhQd] 2006½ okys l engkae DyLVj
djrs gš vKš bl s Bimbam ¼ fož , M LVhQd] 2007½
uked , d , l kš l , 'ku&ijh{k.k dk; Øe eaykxwfd; k x; k
gš ; g e, My l kēU; gšyK/kbll dk čfrfuf/kRo djuš ds
fy, K včdkf'kr voLFkkvkšds, d l š dksfufnžV djrk
gšvKš čR; d DyLVj ¼th½ dks l kbV / ij čR; d DyLVj
eafufgr gšyK/kbll dšvāk dšvuijkr ea, d otu ¼_k½
l kš k tkrk gš

$$\sum_k a_{kl} = 1$$

çR; d DyLVj eaçR; d l kbV ij , yhy 1 dh vkofÜk
 ga FastPHASE e, My tul ç; k gsyk/bi voLFkkvks ds
 chp l Øe.k ds l Fk , d HMM ds: i ea

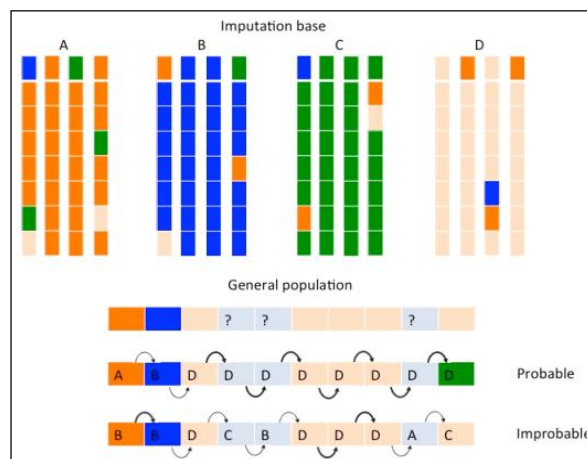
$$P(G_i | \alpha, \theta, r) = \sum_z P(G_i | Z_i, \theta) P | \alpha, r)$$

Pr(G₁|Z,Θ) if jHkkf'kr djrk gSfd n[ksx, thukS/kbi
vltS voLFkk dschp flop djusdse,My iS/uzdh fdruh
l Hkkouk gStksl engka dk cfrfuf/KRo djrsgu fd l anHkz
gSyks/kbiA vltS ;gk] ,d l Hkkouk ds : i eafy[kk tk
l drk gA

$$L(\mathbf{G}; \mathbf{G} \mid \alpha, \theta, r) = \prod P(\mathbf{G}_i \mid \alpha, \theta, r) \prod P(\mathbf{H}_i \mid \alpha, \theta, r)$$

e,My dksfQV djustsfy, ,d bž e ,Yxksjne dk
mi ;ks fd ;k tkrk g\$ v\$ i\$kehVj vupeku ij l'krZ
fehl x thuk/bki dksyxwdjustsfy, ,d QjOM&sdOMZ
,Yxksjne dk mi ;ks fd ;k tkrk g\$ 'kkkdrkZka us i k ;k
fd vupekukads, d l \$ dsvk r l s d o y , d vupeku dh
ryuk ea cgrj cgrj ifj .kke feyA FastPHASE LV\$ ds
Nk/s l \$/ka dk mi ;ks djrk g\$ tks fd ; Fkkspr : i l s
dEl ;W\$ kuy : i l s rst g\$uk pkfg,] gkykd] ; g ykHk
vka'kd : i l s veirZl egkads l kFk dke djustsfv/kd g\$rk
g\$ftl dsfy, dbž ekinb/ka dh vko' ; drk g\$rh g\$

tʃ k f d f p= 2 e a nʹkʲ k s x; k gʃ gʲyʲkʲ b i d k s
DyLVj ekuk tʃr k gʲ a , j c h j l h v kʲ Mh ek d k b J a k y k d s
f o f h k u l i l k a d k s i f j h k kʲ r d j r s gʲ t k s gʲyʲkʲ b i f o r j . k



fp= 2%0,t d 12015½ ij vk/kfjr QkLVQst dsI kFk
thuk/bji bh: W\$kuA

mRi lu d j r s g a , y h y j a k & d k M r g k r s g i v k s v y x & v y x
j a k f d l h f n , x , L F k k u i j l e k u , y h y d s v u q i g k r s
g i v k s l k f k g h l F a s t P H A S E e a l I M P U T E d h r y u k e a l l k k o u k
x . k u k e a v f / k d v k k r i s k e h V j ' k k f e y g s l k f k g h D y L V j
o t u d s c H k k o l F a s t P H A S E e a i s k e h V j v u e k u k a i j v f / k d
e k u d = q v ; k a g a p j . k c) b E l ; q ' k u v k / k k j l s c k l r e w ; k a
i j d n e k i n a k a d k s r ; d j u k , d l l k k f o r l e k / k k u g s
l d r k g s l e k f p z u h , M q k o h l 2010/4

efb $\frac{1}{2}$ Mach $\frac{1}{2}$

MaCH , d HMM e, My dks fu; kft r djrk gS tks fd
IMPUTE dS l eku g\$ yfdU] IMPUTE v\$ FastPHASE dS
foi jhr] MaCH dks, d vyx bti; W\$ku d dh vko'; drk
ughag\$ tcf d bl eor\$ku vu\$pkuk dS vk/kkj ij çR; d
0; fä dS thuk\$ kbi M\$/k dks pj .kc) vi M\$/ djdS vU;
l Hkh uewuk dks vu\$pkfur fd; k tkrk g\$ bl e, My dks bl
çdkj fy [kk tk l drk g\$

$$P(G_i | D_i, \theta, \eta) = \sum P(G_i | Z, \eta) P(Z | D_i, \theta)$$

D.] dks NkM-dj vuekfur g\$yK/kbi dk l \$/ g\$
z HMM dh fNih gPz voLFkk g\$ η ; g fu/kkZjr djrk g\$fd
d,ih fd, x, g\$yK/kbi l sfdrus l eku g\$ v\$ θ fNih
gPz voLFkk ds chp l Øe.k dks fu; i=r djrk g\$vekfpuh
v\$ gkoh 2010% l efV g\$yK/kbi forj.k dks i\$ujkoÜk : i
l sfu/kkZjr fd; k tkrk g\$ v\$ çR; d i\$ujkofÜk ds nk\$ku
i\$kehVj η v\$ θ Hkh viM\$ fd, tkrsg\$ g\$yK/kbi l ds
, d l nHkz i\$y H ds vk/kkj i\$ v\$ çR; kr thuk\$ kbi dks

bā; w'sku vuēkfur gsyk/kbi d; ea H dks tkMēj l ek; k'ftr fd; k tkrk gā fQj gsyk/kbi uemakdsi ujkoLk djdsvcfrēi/kr thuk/kbi dsl hekr forj.k dk vuēku fd; k tkrk gā

chxy ½BEAGLE½

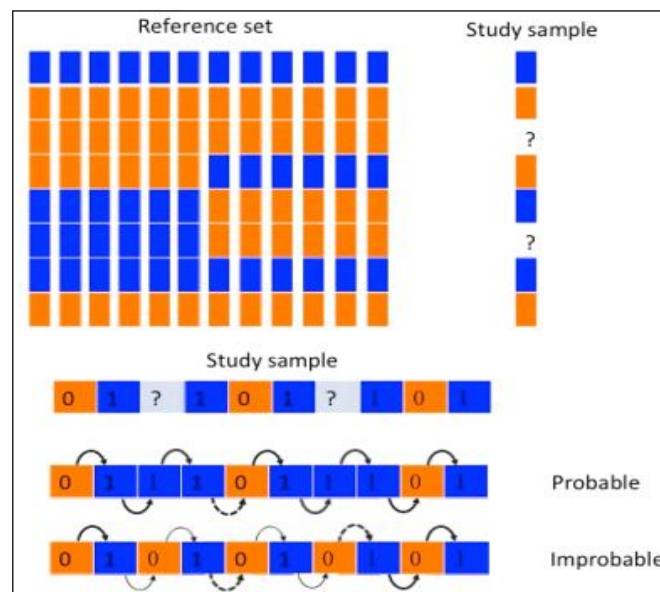
chxy fof/k ½ctmfuax½ 2006½, d xqkl ½ ds l kFk çR; d ekdĳ dh fLFkr ea gsyk/kbi dks LFkkuh; : i l sl emg cukdj, d HMM cukrh gā l emgkads, yhy ij ifjHkkf"kr fd; k x; k gs u fd gsyk/kbi Lrj ijA FastPHASE ds l kFk, d vĳ vrj; g gsfd l emgkadh l ĳ; k fuf'pr ugha gksh çfyd {k=&fuHĳ} gā bl dse, My ea dkbz iĳkehVj ughaGsfTl dk vuēku yxkusdh vko'; drk gsvĳ bl s nkspj. kkaeagsyk/kbi dsfn, x, l ½ ij ykxwfd; k tkrk gā of'od gsyk/kbi vkofĳk; ka dsctk; l ehi LFk ekdĳ ka dschp LFkkuh; Lrj ij, yhy, l kēl, 'ku l Øe.k dh l Hkkoukvka dks fu/kkĳr djrk gā oākkuØe i q l ā kstu çfØ; k dkse, My djusdh dkbz vko'; drk ughaGSD; kēd ekdkb Jĳkyk ea dkbz nh? kdkfyu Lefr ughaGĳ vĳ fdl h Hkh mRi fforĳ dh vuēfr nusdh vko'; drk ughaGksh gā D; kēd chxy eaekdkb Jĳkyk ds, dek= voLFk okLro ea nĳksx, , yhy gā; sl Hkh fo'kkrk, aĳij crk, x, vl; rjhdka dh rĳuk ea thuk/kbi bā; w'sku ea BEAGLE dks dEI; w'skuy : i l srst cukrh gā gkykfd] nkĳk; g gsfd BEAGLE dks i wĳ v/; ; u uemus ij vk/kfĳr gkuk pkfg,]

ftl dk vFkZgsfd bl s IMPUTE2 dh rĳg foHkkftr ugha fd; k tk l drk gā; fn i Mē h ekdĳ ka ij, d çfĳl kc, yhy l ā kstu okysfdl h Hkh gsyk/kbi dks Lopkfyr : i l s'kē; dh l Hkkouk l kē h tk, xhA

chxy fof/k gsyk/kbi Lrj dsctk; , yhy ij l emgk dks ifjHkkf"kr djrh gĳ, l, uih dsfy, l emgkadh l ĳ; k çR; d fLFkr ½uhyk; k ukjāh½ ea nks dsçkçj gksh gā bā; w'sku d vĳ LVMh l ā y ea bā/j&ekdĳ, yhy, l kēl, 'ku dsnkulaLFkkuh; Lrj gsyk/kbi forj.k l Hkkoukvka dks fu/kkĳr djrs gā Vwsgg rhj e/; orhĳl Øe.k l Hkkoukvka dks bāx djrs gā ½fp= 3½A

bEI; w'sku fof/k; kadschp rĳuk

'kkskdrĳka us bu vkerĳ ij bLrēky fd, tkus okys thuk/kbi bā; w'sku Yēodĳĳekfĳh, M gkōĳ 2009; ukĳkuxy, V vy] 2009; okax, V vy] 2012; fy; w, V vy] 2014½ dsçn'kē dk l ko/kkuhi wĳd v/; ; u fd; k gā ekfĳh, M gkōh ½2010½ us l nHĳ i ā y ds vuĳ kj l cl s ykdfç; bā; w'sku fof/k; ka ea l s çR; d ds xqkka dks l ā ki ea çLrĳ fd; k gĳtkĳrkfydk 1½ v/; ; u uemakdsxqk ½rkfydk 2½ dk; Øe fodYi ½rkfydk 3½ dEI; w'skuy çn'kē vĳ =qV nj ½rkfydk 4½A mĳgkūs ik; k fd IMPUTEv1, MACH, FastPHASE vĳ BEAGLE dschp IMPUTEv2 mudsfl ey'sku ifj-'; vĳ, d MAR/kkĳ.kk ds vk/kkĳ ij l cl sl Vhd -f"Vdksk gĳ yfdu l Hkh fof/k; k; l eku çn'kē mRi l u



fp= 3%chxy dsl kFk thuk/kbi bā; w'sku Ø,td ½2015½ ij vk/kfĳr FastPHASE l s vyXA

djrh gā bl ds vykokl vks dh tkp djus ds fy, fd 1000 thukē cktōV tš sgšykb/kl dscMā l nHkz i šy ij rjhd sš sčn'kū djrsgā 'kkdkrkz/kaus 1000 thukē cktōV l sik; yV l hbz wgsykb/kl dsvk/kkj ij 500 vks 1]000 0; fā; kāl s; qā 1]000 gšykb/kl dsl nHkz i šy dk mi; kx fd; k gšvks HAPGEN fl eysku ykxwfd; k gā xqkl = 10 ij 5, ech {ks=A ifj.kke fn[kkrsgā d IMPUTEv2] BEAGLE vks FastPHASE dh rgyuk earst gā

okx, V vy ½2012½ us, d ubz fudVre i Mh h fof/k ¼ u, u½ vks, d Hkktjr l d j.k ¼ MCY; wu, u½ dk çLrko j [kk] tksnkuka l gl a kst d fl) kr dk ikyu djrsgā yf{kr 0; fā ds ikl tul ē; k l s, d ds l eku thukv/kl vupe

glrk gā bu nks fof/k; k ds vykokl mlglkus FastPHASE, Npute ½j, CV½, V vyA] 2007½ vks dbz e'khu yfuž bā; wš ku fof/k; k dks Hkh ykxwfd; k] ft l ea l i kšZ ošVj e'khu ½svm½, d LFkkuh; U; j y us/odZ½; j y us½ vks, d LFkkuh; Ql ½ v, Mj ekdkb pu ¼mc½ 'kkfey gā ifj.kkeka l s i rk pyk fd, u, u vks MCY; wu, u l cl s d qky rjhdka ea l s Fkš vks ykirk, l, uih thukv/kl bā; wš ku ea QKLVO d dks NkM elj vU; rjhdka dh rgyuk ea dk Qh cgrj čn'kū fd; kA

fy; w, V vy ½2014½ 90 0; fā; kāl s i j & thukē Mh, u, vupe. k Mš/k dsvk/kkj ij thukv/kl bā; wš ku čn'kū dh 0; ofLFkr : i l stkp dhA çfr: i. k čn'kū dk eV; kadu

rkfydk 1% l nHkz xqkA

xqk	bEl; wš ku fof/k				
	IMPUTEv1	IMPUTEv2-2	MACHv1-0-16	FastPHASEv1-4-0	BEAGLEv3-2
l nHkz i šy					
gšykb/kl l nHkz i šy dk mi; kx dj l drsgā	gk	gk	gk	gk	gk
thukv/kl M jōjā i šy dk mi; kx dj l drsgā	ugha	gk	gk	gk	gk
D; k nks gšykb/kl ; k thukv/kl M jōjā i šy, d gh ju ea blrky fd, tk l drsgā	ugha	gk	ugha	ugha	ugha
l nHkz i šy l gh čk: i ea miyC/k gā	HapMap2 HapMap3 IKGP pilot data	HapMap2 HapMap3 IKGP pilot data 1000 Genome projects	HapMap2 HapMap3 IKGP pilot data 1000 Genome projects	Hap2	1000 Genome projects

rkfydk 2% v/; ; u ds ueus ds xqkA

xqk	bEl; wš ku fof/k				
	IMPUTEv1	IMPUTEv2-2	MACHv1-0-16	FastPHASEv1-4-0	BEAGLEv3-2
v/; ; u ds ueus					
vfuf' prrk ds l kFk fufnZV thukv/kl ys l drsgā	ugha	gk	ugha	ugha	gk
frdMh vks l cā/kr ueuka dks l ek; kštr dj l drsgā	ugha	ugha	ugha	ugha	gk
D; k v, Vkl key gšykb/kl dk, d v/; ; u ueuk yxk; k tk l drk gā	gk	gk	ugha	ugha	gk
x xqkl = ij bEl; w dj l drsgā	gk	gk	ugha	ugha	gk

djusdsfy, ekunM ds: i eal Vhd : i l syxk, x, døy l mHKz i sy dk mi; kx djus dh ryuk ea cgrj
 ofj, v ds cfr'kr dk mi; kx djd l cl sigy mlglus çn'kz fn[kk; k gA, d gh vkcknhA nll jk] tkpdrkzvkerk
 ik; k fd fefuef vls IMPUTE2 ea BEAGLE dh ryuk ea ij vksdsfo'ySk.k l s[kjk : i l syxk, x, ofj, v dks
 cgrj bñ; W\$ku çn'kz gsvls cgrtutl f; k l mHKz i sy us gvksdsfy, bñ; W\$ku DokfyVh eki ij Hkjkd k djsrgA

rkfydk 3%dk; Døe fodYi ka vls I fo/kkvka ds xqkA

xqk	bEl; W\$ku fof/k				
	IMPUTEv1	IMPUTEv2-2	MACHv1-0-16	FastPHASEv1-4-0	BEAGLEv3-2
çkxke dsfodYi vls fo'kkrk, a					
D; k QfI æ vls I kFk gh	ugha	gk	gk	gk	gk
bEl; W\$ku mi yC/k gS\					
D; k vkrfjd çn'kz eW; kadu gS	gk	gk	gk	ugha	ugha
døy, d fufnZV varjky ea	gk	gk	ugha	ugha	ugha
gh bEl; W dj I drsgA					
MV/k I v ds chp LVMM I j f k.k	gk	gk	gk	ugha	ugha
dk l Hkky I drk gS					
, l , uih vls ueuk I eko\$ku	gk	gk	ugha	gk	gk
, oa cfg"dj.k fodYi \					
bñ; W\$ku vls , l kEl , 'ku ij h{k.k	ugha	ugha	ugha	ugha	ugha
dsfy, I a e, My\					
v, ij fVx fl LVe dh vko'; drk\	Linux, Solaris, Windows, Mac	Linux, Solaris, Windows, Mac	Linux, Windows, Mac	Linux, Solaris, Windows, Mac	Java executable

rkfydk 4%del; W\$ku y çn'kzA

xqk	bEl; W\$ku fof/k				
	IMPUTEv1	IMPUTEv2-2	MACHv1-0-16	FastPHASEv1-4-0	BEAGLEv3-2
del; W\$ku y çn'kz					
vkdyu 1*	ugha	gk	ugha	ugha	gk
vkdyu 2*	ugha	ugha	ugha	ugha	gk
=fV njç					
i ä; k ifj-'; A ds vuq i gS	5-42%	5-16%	5-46%	5-92%	6-33%
ifj-'; ch %çfrcA/kr½	---	3-40%	---	5-33%	3-46%
ifj-'; ch %i wlk½	---	3-40%	---		4-01%

★120 CEU Hapmap2 haplotypes l s Affy500k fpi ij 1377 ueukadk bEl; W\$ku; 7-5, ech {k= A MV/k %gkoh}, V vya] 2009%
 l svkrk gA # 500 %1000% ueukadk bEl; W\$ku 5, ech {k= ea 8712, l , uih ij 872, l , uih l s 872, l , uih ij thukv/bi
 fd; k x; kA Økæd k 10 ij 5, ech {k= ea 1000 thuke ifj; kstuk l SHAPGEN vls ik; yV CEU gsyk/bi dk mi; kx djd
 fl E; gV/M MV/k l v ij vk/fjr l e; A ç=fV nj %gkoh}, V vya] 2009% IMPUTE2 ds ifj.kke vi MV dj fn, x, gA fn,
 x, ifj-'; B =fV nj Affymetrix SNPs l syxk, X, Illumina SNPs dsfy, gA Illumina, l , uih l syxk, x, , fQe fV DI
 , l , uih dsfy, =fV nj çbV eanh xbZ gA

fu"d"l

bl v/; ; uea fopkj fd, x, febl x Mv/k ifj-' ; ka ea IMPUTEv2 eamPp l Vhdkr vls rst x.kuk nksukaFka gkoh , V vy 1/2011½ dk ekuuk gSfd IMPUTE2 dh l Qyrk bl dh dEl; w/skuy j.kuhfr; ka vls vius Mh, u, vupe fHKUrk dse, My dh otg l sgA e, My ds, YxksjFe ds -f"Vdksk l } BEAGLE vls FastPHASE gSyk/kbi dksl euka ea tkMfsgs tks x.kuk cfØ; k dks xfr nrs gSD; kcd ; g HMM voLFkkvka dh l f; k dks l hfer djrk gA BEAGLE vls FastPHASE dksMvkl v eaçR; d gSyk/kbi ij HMM x.kuk djus ds ctk; doy l euka ds, d Nks l v ij x.kuk pykus dh vko'; drk gsrh gA bl ds foijhr] IMPUTEv2 vls MaCH, d LV/ ea çR; d gSyk/kbi ds fy, HMM dk çn'ku djrs gS gkykad] l Hkh LV/ dk mi ; kx x.kuk dksdfBu cuk nrk gS vls bl fy, IMPUTEv2 jkT; ka voLFkkvka dks çfrca/kr djrk gA

gkoh , V vy 1/2011½ usçnf'kr fd; k fd IMPUTEv2 muds } jk tks x, ifj-' ; kadsfy, BEAGLE dh ryuk eamPp l Vhdkr çkr djrk gS vls ; g Mvkl v eade&vkoflk okysofj, v eafoksk : i l sLi"V gSftl eamPp gSyk/kbi fofo/krk gA thuk/kbi bh; w/sku ftl ustfvy ekuo jkska dks l dMka okLrfod l akka rd igpk; k gS thuke&okBM , l ksl , 'ku LVMht 1/6WAS½ dk , d vko'; d fgLi k cu jgk gS fgmksQ, V vy] 2009% vls ; g vxysdN o"kkard tkjh jgskA eq; dkjd tskchkfor djrk fd fd l bh; w/sku fof/k dk mi ; kx fd; k tkrk gS osgks tks vxyh ih< ds vupe.k Mv/k dh c<rh mi yC/krk dksMh l f; k ea, l , uih dsl kfk l Hky l drsgA gSyk/kbi dscMh vf/kd fofo/k l v dk mi ; kx djuseavkji .k fof/k; kadsfy, Hkh ; gh pqlsh gA

vr ep thuk/kbi bh; w/sku dksç; kx'kkyk&vk/kfjr Mv/k ih< tS svPNs okkfud vH; kl dsl eku fu; eka dk ikyu djuk pkfg, A vkupf'kd jks l ak v/; ; u 1/4 Mj l u , V vy] 2010½ ea Mv/k xqkoUk ds ekunM dks ifjHkkf'kr djusdsfy, vrhr eacgr l kjsç; kl fd, x, gA bl fy, thuk/kbi bh; w/sku dsfy, l eku Lrj dh l Vhdkr vls fo'ol uh; rk eW; kadu ekunM fodfl r djuk vko'; d gA

I UnH

Anderson, C.A., Pettersson, F.H. and Clarke, G.M. (2010). Data quality control in genetic case-control association studies. *Nature Protocols*. 5: 1564-1573.

Browning, S.R. (2006). Multilocus association mapping using variable-length Markov chains. *American Journal of Human Genetics*. 78: 273-280.

Cheema, J.R. (2014). A review of missing data handling methods in educational research. *Review of Educational Research*. XX(X): 1-22.

Fearnhead, P. and Donnelly, P. (2001). Estimating recombination rates from population genetic data. *Genetics*. 159: 1299-1318.

Hindorff, L.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*. 106(23): 9362-9367.

Howie, B. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*. 44(8): 955-960.

Howie, B., Marchini, J. and Stephens. (2011). Genotype imputation with thousands of genomes. *G3: Genes, Genomes, Genetics*. 1: 457-469.

Howie, B.N., Donnelly, P. and Marchini, J. (2009). A flexible and accurate method for the next generation of genome-wide association studies. *Plos Genetics*. 5(6): e1000529. doi:10.1371/journal.pgen.1000529.

Krawczak, M. (2015). Genotype Imputation. In: eLS. John Wiley and Sons, Ltd: Chichester. doi: 10.1002/9780470015902.a0022399.

Li, N. and Stephens, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*. 165: 2213-2233.

Liu, Q. (2014). Systematic assessment of imputation performance using the 1000 Genomes reference panels. *Briefings in Bioinformatics*. DOI: 10.1093/bib/bbu035.

Marchini, J. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics*. 39(7): 906-913.

Marchini, J. and Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*. 11: 499-511.

Nothnagel, M. (2009). A comprehensive evaluation of SNP genotype imputation. *Human Genetics*. 125: 163-171.

Roberts, A. (2007). Inferring missing genotypes in large SNP panels using fast nearest-neighbor searches over sliding windows. *Bioinformatics*. 23: 401-407.

Roth, P. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*. 47: 537-560. doi: 10.1111/j.1744-6570.1994.tb01738.x.

Rubin, D.B. (1976). Inference and missing data. *Biometrika*. 63(3): 581-592.

Servin, B. and Stephens, M. (2007). Imputation-based analysis of association studies: Candidate Regions and Qualitative Traits. *Plos Genetics*. 3(7): e114. doi: 10.1371/journal.pgen. 0030114.

Sheet, P. and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *The American Journal of Human Genetics*. 78: 629-644.

Wang, Y.N., Cai, Z.P., Stothard, P. (2012). Fast accurate missing SNP genotype local imputation. *BMC Research Notes*. 5: 404. doi: 10.1186/1756-0500-5-404.