LB-Impute is designed to impute low-coverage sequencing datasets generated from biallelic populations. It is designed to work on version 4.1 VCF files that contain allelic depth of coverage information, designated by the field AD. At the present time, LB-Impute only works in the command line, though a GUI version will be introduced soon. LB-Impute has been tested on Mac OS 10.9.5 and Red Hat Enterprise Linux Server release 6.2 (Santiago). LB-Impute was compiled under Java version 1.6 (Java 6).

Parental sequencing:

For LB-Impute to work, some parental sequencing is required. Parental samples should be in the imputed VCF file. Parents should be filtered so that only markers that are homozygous within and polymorphic between are left in the parental dataset. Stringent filtering on other criteria is also recommended. LB-Impute can perform a parental imputation step to determine the remaining, ambiguous markers with high accuracy, even if most sites are not called in the parents.

Output

The output of LB-Impute is a new VCF file, with all imputed values substituted for the genotype plus 99 in every other field. Non-imputed values will be set to ./. to indicate they are missing. Note that LB-Impute attempts to impute every genotype, including ones that were previously called, and may remove existing genotypes from the dataset if it finds them ambiguous.

Example Commands:

IMPUTE

java -jar LB-Impute.jar -method impute -f vcffile.vcf -readerr 0.05 -genotypeerr 0.05 -recombdist 10000000 -window 5 -offspringimpute -parents par1,par2 -o outputvcf.vcf

For use with the example files, this would be:

java -jar LB-Impute.jar -method impute -f example_data/simulated_data/F1BC1/Coveragemodifiedvalues_chrl100000000_scount200_ptypef1bc1_vcount10000_mrec3_stdrec2_reads5000_rep1.vcf -readerr 0.05

-genotypeerr 0.05 -recombdist 10000000 -window 5 -offspringimpute
-parents par1,par2 -o outputvcf.vcf

or

java -Xmx3g -jar LB-Impute.jar -method impute -f
example_data/GBS_data/HincII_7rds_0.5_17436removed_872501kept.vcf -readerr
0.05 -genotypeerr 0.05 -recombdist 10000000 -window 5 -offspringimpute -parents
HincII_B73,HincII_CG -o outputvcf.vcf

COMPARE

java -jar LB-Impute.jar -method compare -missfile validationset.vcf -imputefile
imputationresults.vcf -mainfile completevcf.vcf -parents par1,par2 -compareoffspring >
Imputationresults.txt

For use with the example files, this would be:

java -jar LB-Impute.jar -method compare -missfile
example_data/simulated_data/F1BC1/Coveragemodifiedvalues_chrl100000000_scount20
0_ptypef1bc1_vcount10000_mrec3_stdrec2_reads5000_rep1.vcf -imputefile
example_data/simulated_data/F1BC1/Imputed_chrl100000000_scount200_ptypef1bc1_v
count10000_mrec3_stdrec2_reads5000_rep1.vcf -mainfile
example_data/simulated_data/F1BC1/Actualvalues_chrl100000000_scount200_ptypef1b
c1_vcount10000_mrec3_stdrec2_reads5000_rep1.vcf -parents par1,par2
-compareoffspring > Imputationresults.txt

or

java -Xmx3g -jar LB-Impute.jar -method compare -missfile
example_data/GBS_data/HincII_7rds_0.5_17436removed_872501kept.vcf -
imputefile
example_data/GBS_data/Imputed_HincII_7rds_0.5_17436removed_872501kept.vcf -
mainfile example_data/GBS_data/HincII_FilteredVariants_Cleaned2.vcf -parents
HincII_B73,HincII_CG -compareoffspring > Imputationresults.txt

RANDOM REMOVE (create validation set)

java -jar LB-Impute.jar -method randomremove -f vcffile.vcf -minhomcov 7
-maxhomcov 7 -minhetcov 7 -maxhetcov 7 -parents par1,par2 -removefraction 0.5

For use with the example files, this would be:
java -jar LB-Impute.jar -method randomremove -f
example_data/simulated_data/F1BC1/Actualvalues_chrl100000000_scount200_ptypef1b
c1_vcount10000_mrec3_stdrec2_reads5000_rep1.vcf -minhomcov 7 -maxhomcov 7

-minhetcov 7 -maxhetcov 7 -parents par1,par2 -removefraction 0.5

or

java -Xmx3g -jar LB-Impute.jar -method randomremove -f example_data/GBS_data/HincII_FilteredVariants_Cleaned2.vcf -minhomcov 7 -maxhomcov 7 -minhetcov 7 -maxhetcov 7 -parents HincII_B73,HincII_CG -removefraction 0.5

Commands

-method	<impute, compare, randomremove>	Determines what operation LB-Impute will undertake. REQUIRED.
IMPUTE OPTIONS		
-f	<filename>	VCF input file name. REQUIRED.
-o	<filename>	Output file name. REQUIRED.
-parents	<parentnames>	Names of parental samples separated by comma. NO SPACE BETWEEN COMMA AND NAMES. Names with spaces or commas will not work properly. REQUIRED.
-offspringimpute		LB-Impute will impute offspring.
-parentimpute		Default. LB-Impute will impute parents.
-readerr	<double>	Default value is 0.05. Probability that any given read is erroneous.
-genotypeerr	<double>	Default value is 0.05. Probability that any given genotype call is erroneous.
-recombdist	<integer>	Default value is 10,000,000. Expected distance of 50cM.

-resolveconflicts		Default is off. When on, it will use the path with the highest probability to resolve a conflict rather than leaving a conflicting genotype empty. May reduce accuracy.
-dr		Default is off. When on, a homozygous to homozygous recombination event (double recombination) will have the same probability as a single event. This may be useful for inbred populations such as NAMs.
-minsamples	<integer>	Default is 5. Minimum number of imputed samples required to infer a parental genotype. Has no function when imputing offspring.
-minfraction	<double>	Default is 0.5. For a genotype at a given locus to be assigned to a parent, that genotype must be at least this fraction of the total genotypes of offspring assigned to that parent. This value should be somewhat low for low coverage populations, as false homozygotes will be confounding.
-window	<integer>	Trellis window length. Default value is 7.

RANDOMREMOVE OPTIONS

Random remove removes calls at a specific level or range of coverage to serve as a validation set. When used along with compare, it can validate imputation on a dataset. Prints new dataset to stdout.

-f	<filename>	Name of VCF file. REQUIRED.
-parents	<parentnames>	Names of parental samples separated by comma. NO SPACE BETWEEN COMMA AND NAMES. Names with spaces or commas will not work properly. REQUIRED.
-removefraction	<double>	Fraction of calls fitting criteria to be removed.
-minhomcov	<int>	Minimum coverage of homozygous calls to be removed. Inclusive. REQUIRED.
-maxhomcov	<int>	Maximum coverage of homozygous calls to be removed. Inclusive. REQUIRED.
-minhetcov	<int>	Minimum coverage of heterozygous calls to be removed. Inclusive. REQUIRED.
-maxhetcov	<int>	Maximum coverage of heterozygous calls to be removed. Inclusive. REQUIRED.

COMPARE OPTIONS

Compare compares the original file with an imputed file and a file with a validation set removed then outputs statistics to stdout.

-mainfile	<filename>	Name of original file. REQUIRED.
-imputefile	<filename>	Name of imputed file. REQUIRED.

-missfile	<filename>	Name of file with validation set removed. REQUIRED.
-offspringcompare		Compares offspring rather than parental calls.
-parents	<parentnames>	Names of parental samples separated by comma. NO SPACE BETWEEN COMMA AND NAMES. Names with spaces or commas will not work properly. REQUIRED.