SPANDx

User Manual

Version 2.1, last modified 13Dec14

For the latest version of SPANDx and the user manual, please visit the website:
http://sourceforge.net/projects/spandx/

CONTENTS

INTRODUCTION AND DESCRIPTION

SPANDx (**S**ynergised **P**ipeline for **A**nalysis of **N**ext-generation sequencing **D**ata in Linu**x**) is a comparative genomics pipeline designed to greatly simplify the identification of genetic variants (SNPs, insertions/deletions (indels), and large (>200bp) deletions) from medium- to large-sized haploid next-generation re-sequencing (NGS) datasets. SPANDx can currently process several NGS data inputs including paired- and single-end data from the Illumina MiSeq, HiSeq and GA$_{IIx}$ platforms, single-end data from the Life Technologies Ion Personal Genome Machine (PGM)®, and single-end Roche 454 data. SPANDx integrates the following validated bioinformatics tools for start-to-finish sequence analysis of raw NGS data using a single command:

- **Burrows-Wheeler Aligner (BWA)** (Li & Durbin, 2009 *Bioinformatics* 25(14):1754-60; Li & Durbin, 2010 *Bioinformatics* 26(5):589-95) for alignment of short (i.e. Illumina and PGM) or long (i.e. 454) NGS read data. BWA is downloadable from: http://bio-bwa.sourceforge.net/. SPANDx does not currently support BWA versions later than 0.6.2.
- **SAMtools** (Li *et al*., 2009 *Bioinformatics* 25(16):2078-79) and **Picard** (as-yet-unpublished) for alignment manipulation and filtering. These programs can be downloaded from http://samtools.sourceforge.net/ and http://picard.sourceforge.net/, respectively.
- **The Genome Analysis Tool Kit (the GATK)** (McKenna *et al*., 2010 *Genome Research* 20 (9):1297-1303; DePristo *et al.,* 2011 *Nat. Genet.* 43(491-98); Van der Auwera *et al.,* 2013 *Curr Prot Bioinform.* 43(11.10.1-11.10.33)) for base quality score recalibration, realignment of regions with low mapping quality, duplicate removal, identification of SNPs and indels, and

filtering the variant call format (VCF) file generated from the alignment process. The GATK can be downloaded from http://www.broadinstitute.org/gatk/.

- **VCFtools** (Danecek *et al*., 2011 *Bioinformatics* 27(15):2156-58) for manipulation of VCF files. VCFtools is downloadable from: http://vcftools.sourceforge.net/. All file outputs from SPANDx are in standardised VCFv4.1 format. **tabix** and **bgzip** are VCFtools dependences that are required for data handling of `.vcf` files, and can be downloaded here: http://sourceforge.net/projects/samtools/files/tabix/.
- **BEDTools** (Quinlan & Hall, 2010 *Bioinformatics* 26(6):841-42), and specifically the *coverageBED* module, for identification of locus presence/absence across each genome of interest based on the reference sequence. This tool is useful for identifying larger-scale (approx. 200bp or larger) deletions. BEDTools can be downloaded from: https://github.com/arq5x/bedtools2/.
- **SnpEff** (Cingolani *et al*., 2012 *Fly* 6(2):80-92) for annotation of SNP and indel variants. SnpEff can be downloaded from here: http://snpeff.sourceforge.net/.

Prior to running SPANDx, these programs need to be installed by the user and the path to their locations specified in the `SPANDx.config` file. The full version of SPANDx will install all above components except for the GATK, which needs to be downloaded and installed separately due to Broad Institute licencing restrictions.

Novel comparative genomic features of SPANDx include:

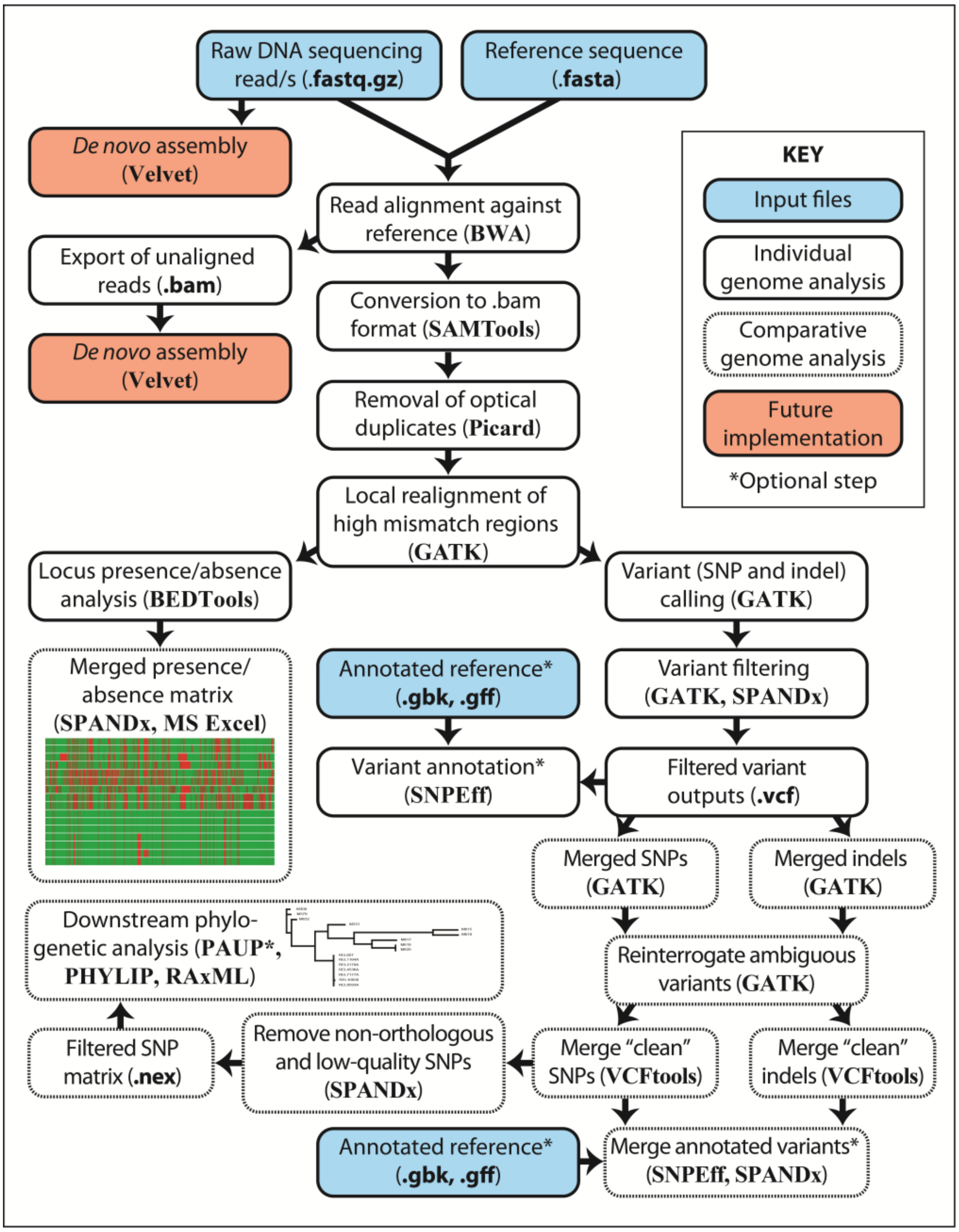
- Merged orthologous[1] SNP and indel matrices that greatly simplify variant visualisation for comparative genomic analyses.
- PAUP*, PHYLIP and RAxML-compatible .nex orthologous[1] SNP matrix files for downstream phylogenetic analyses. PHYLIP and RAxML are freely available programs downloadable from: http://evolution.genetics.washington.edu/phylip.html and http://www.exelixis-lab.org/.
- Locus presence/absence matrices from BEDTools outputs that enable simple visualisation and comparative genomic determination of 1) the 'core' genome and 2) variable genetic loci (including deleted regions brought about by reductive evolution).
- Merged, annotated SNP and indel matrices for fast and simple genetic characterisation of variants (NB. The user must provide SPANDx with the SnpEff-annotated reference genome information for variant annotation to work).

[1]The term "orthologous" refers to genetic loci that make up the 'core' genome. SNPs or indels residing in genetic loci that are missing in one or more genomes are excluded. If these variants are required, they can be found in the individual filtered .vcf output files.

Unlike many other bioinformatics tools, SPANDx does not require pre-assembled genomes, is written in non-compiled code and is customisable. In addition, the default settings for variant calling using Illumina, Ion PGM and 454 data have already been optimised and do not require the user to specify these settings, although these settings can be customised if required.

SPANDx has been purposely written for systems that utilise Portable Batch System (PBS) or Sun Grid Engine (SGE) [i.e. qsub] to enable task parallelisation, greatly reducing turn-around-time of datasets comprising tens through to thousands of haploid genomes using a single command. Currently there is no support for SPANDx on other resource management systems (e.g. SLURM, LSF) due to the unavailability of such systems in our laboratory, but compatibility with resource managers can be addressed if required. Please email us at mshr.bioinformatics@gmail.com if you require a specific resource manager compatible version of SPANDx and are willing to test it on your system.

The SPANDx workflow is shown below:


Raw DNA sequencing read/s (.fastq.gz)
Reference sequence (.fasta)
De novo assembly (Velvet)
Read alignment against reference (BWA)
Export of unaligned reads (.bam)
Conversion to .bam format (SAMTools)
De novo assembly (Velvet)
Removal of optical duplicates (Picard)
Local realignment of high mismatch regions (GATK)
Locus presence/absence analysis (BEDTools)
Variant (SNP and indel) calling (GATK)
Merged presence/ absence matrix (SPANDx, MS Excel)
Annotated reference* (.gbk, .gff)
Variant filtering (GATK, SPANDx)
Variant annotation* (SNPEff)
Filtered variant outputs (.vcf)
Merged SNPs (GATK)
Merged indels (GATK)
Downstream phylo-genetic analysis (PAUP*, PHYLIP, RAxML)
Reinterrogate ambiguous variants (GATK)
Filtered SNP matrix (.nex)
Remove non-orthologous and low-quality SNPs (SPANDx)
Merge "clean" SNPs (VCFtools)
Merge "clean" indels (VCFtools)
Annotated reference* (.gbk, .gff)
Merge annotated variants* (SNPEff, SPANDx)
KEY
Input files
Individual genome analysis
Comparative genome analysis
Future implementation
*Optional step

SYNOPSIS

```
SPANDx -r <exact reference name, excluding .fasta extension> [parameters,
optional] -o [organism] -m [generate SNP matrix yes/no] -i [generate indel
matrix yes/no] -a [include annotation yes/no] -v [reference file for
variant annotation; name must exactly match the SnpEff database name, which
can be found in the snpEff.config file] -s [specify read prefix to run
single strain; set to none to construct a SNP matrix from a previous
analysis or leave as default to process all reads] -t [sequencing platform
i.e. Illumina/Illumina_old/454/PGM] -p [pairing of reads (i.e. paired-end
or single-end) PE/SE] -w [BEDTools window size in base pairs]
```

COMMANDS AND OPTIONS

`SPANDx.sh` is the only script that needs to be run to obtain data outputs. SPANDx by default expects paired-end Illumina data with v1.8+ quality encoding. If your data are in this format, the only required switch is `-r` to specify the reference sequence prefix. If another NGS data format is to be analysed, please specify this format using the `-t` (and if single-end, the `-p`) switch/es[2]. By default, SPANDx *will* construct a locus presence/absence matrix but *will not* construct orthologous SNP or indel matrices, nor will it perform variant annotation. The `-m` (matrix) and `-i` (indel) switches are required for orthologous SNP and indel matrix creation, respectively. The `-a` (annotate) and `-v` (reference name for variant identification) switches are both required for variant annotation. Prior to running `SPANDx.sh`, both the reference (in `.fasta` format) and NGS files (in `.fastq.gz` format) are required to be in your analysis directory. SPANDx expects NGS reads to conform to the following naming format regardless of the sequencing technology used: `strain_1_sequence.fastq.gz` and `strain_2_sequence.fastq.gz` (for paired-end reads) or `strain_1_sequence.fastq.gz` (for single-end reads).

[2] SPANDx cannot process multiple NGS formats (e.g. single-end and paired-end Illumina) in a single run. If multiple NGS formats are to be analysed, please create separate analysis directories and run SPANDx specifically for each NGS format. These data can be merged for downstream analysis. See the `-s` description for more information.

Options:

`-r` *STR* *Required*. Specifies the reference genome file, excluding the `.fasta` extension. The `-r` switch is the only mandatory switch needed for SPANDx to function. Additional switches are required to modify the default behaviour of SPANDx and sequencing technology needs to be specified if your data are not paired-end Illumina data with v1.8+ quality scores. The reference file is required to be in `.fasta` format and should conform to the standard FASTA specification, as detailed here: http://www.ncbi.nlm.nih.gov/BLAST/blastcgihelp.shtml. IUPAC codes are not supported by some programs incorporated in SPANDx and *must* be avoided. For compatibility with the annotation module of SPANDx, the chromosome names for the reference genome must match those used by SnpEff. This nomenclature can be found in the `snpEff.config` file, which is generated upon SnpEff installation, or automatically with the full SPANDx installation. In addition, the `.fasta` reference file must not contain any blank lines.

-o *STR* *Optional.* Specifies the organism under analysis. The -o parameter is used in naming the read group headers in the SAM and BAM files. Spaces and special characters may have unexpected behaviour and should be avoided. [*Haploid*]

-m *yes/no* *Optional*. The -m switch is used to create a matrix with all orthologous SNP variants identified in the analysis. Non-orthologous SNPs are excluded. Output .nex files can be directly imported into PAUP*, PHYLIP or RAxML for phylogenetic analysis. By default this behaviour is switched off. [*no*]

-i *yes/no* *Optional*. The -i switch is used to create a matrix with all orthologous indel variants identified in the analysis. Non-orthologous indels are excluded. By default this behaviour is switched off. [*no*]

-a *yes/no* *Optional*. The -a switch is used to perform annotation of the variant files. By default this behaviour is switched off. If annotation is switched on the -v switch must also be specified. [*no*]

-v *STR* *Optional, required if -a is set to "yes".* The -v switch is used to specify the reference genome that SnpEff will use to annotate variants. The string used for this variable must match one of the genomes contained within the SnpEff.config file. Additionally, chromosome names in the reference file must match those contained within the SnpEff program. Please refer to the SnpEff manual (which can be found here: http://snpeff.sourceforge.net/SnpEff_manual.html) for more information. [*null*]

-s *STR* *Optional*. The -s switch is used to flag a single genome for analysis, or as a special case to create the orthologous SNP matrix from previously generated variant calls and alignments. SPANDx expects all NGS reads to be in the sequence analysis directory (i.e. the present working directory) and by default all NGS reads within the sequence analysis directory will be processed. Before running SPANDx, make sure the reference genome and NGS read files are in the sequence analysis directory and that the NGS read files conform to the following format (strain_1_sequence.fastq.gz and strain_2_sequence.fastq.gz for paired-end reads, or strain_1_sequence.fastq.gz for single-end reads). If -s is set to none, SPANDx will not perform individual analysis of any NGS read files in the current directory. Instead, SPANDx will move to the comparative genomics section of the pipeline (see SPANDx workflow above) and assume all individual genome analysis has already been completed. SPANDx will then merge all VCF files contained in $PWD/Phylo/snps and perform error correction using the .bam and .bai files contained in $PWD/Phylo/bams to construct an orthologous SNP matrix, which will be output to $PWD/Phylo/out. Before running this module, please check that all .vcf files located within $PWD/Phylo/snps match the alignment files within $PWD/Phylo/bams, and that all .bam files contain their accompanying .bai index file. This feature is useful for mitigating the need to re-run previous analyses from scratch, or for combining data generated from multiple SPANDx runs (e.g. from different sequencing technologies) into a single orthologous SNP matrix. [*all*]

-t *STR* *Optional.* The -t switch specifies the sequence technology used, and must be one of the following: Illumina, Illumina_old, 454 or PGM. By default, SPANDx expects Illumina reads with Phred+33 read quality encoding, which is standard as of v1.8+. To specify NGS reads generated by the older Illumina format (i.e. Phred+64), use `-t Illumina_old`. If the analysis mode is switched to `454` or `PGM`, SPANDx will use the BWA-SW algorithm of BWA for read alignment and thus will expect reads to be in single-end `strain_1_sequence.fastq.gz` format with Phred+33 quality encoding. [*Illumina*]

`-p` *STR* *Optional*. The `-p` switch specifies the pairing of reads and must be either `PE` or `SE`. By default SPANDx expects reads to be paired. If reads are single end, `-p` *must* be set to `SE`. Currently SPANDx does not support paired-end 454 or PGM data. [*PE*]

`-w` *INT* Optional. The `-w` switch specifies the window size (in base pairs) used by BEDTools to analyse whole genome alignment coverage i.e. locus presence/absence. [*1000*]

INSTALLATION AND REQUIREMENTS

SPANDx is written in Bash and will run on most Linux installations that have PBS or SGE (i.e. qsub). SPANDx has been tested on GNU Bash version 3.2.25(1)-release and GNU Bash version 4.1.2(1)-release (both on x86_64-redhat-linux-gnu) with Java v1.7.0_55 and v1.7.0_71. The 2.4 version of SPANDx (downloadable here: http://sourceforge.net/projects/spandx/) has been tested using PBS (TORQUE v2.5.13) and SGE v6.2u5p3, BWA v0.6.2, SAMtools v0.1.19, Picard v1.105, the Genome Analysis Toolkit v3.0, BEDTools v2.18.2, SnpEff v3.5, VCFtools v0.1.11 and tabix v0.2.6. SPANDx does not currently support BWA versions later than 0.6.2.

> At this time, SPANDx requires PBS or SGE to submit jobs to the cluster. If you do not have these system setups, SPANDx will not function. Please contact us if you are using a different Linux scheduler setup and are willing to trial-and-error SPANDx on your system.

SPANDx comes in two flavours: lite and full. The full version is easier to install and contains almost all other programs (compiled with x86_64 architecture) utilised within the SPANDx pipeline, with the exception of the GATK. To install the full version of SPANDx, gunzip and untar (usually `tar xvfz SPANDx_full.tar.gz`) the SPANDx full distribution in your /bin directory. **IMPORTANT:** following extraction of the script files, the `SPANDx.config` file will need to be modified to contain the location of the SPANDx installation. If your system uses a proxy to access the internet, please modify the JAVA_PROXY variable in `SPANDx.config`. For default SPANDx behaviour, no other settings should need to be modified. In addition, you will also need to download and install the GATK and either place the GenomeAnalysisTK.jar file (renamed without version numbers) in the SPANDx installation directory or specify the install path within the `SPANDx.config` file.

For users of non x86_64 systems, or those with SPANDx dependencies already installed, the lite version contains the entire SPANDx program but none of the dependencies. For the lite version, gunzip and untar SPANDx (`tar xvfz SPANDx_lite.tar.gz`) in the `/home/user/bin/SPANDx` directory. SPANDx should work from any installation directory but has been most extensively tested in `/home/user/bin`. **IMPORTANT:** following extraction of the script files, the `SPANDx.config` file will need to be modified to contain the absolute paths of each dependency (i.e. BWA, SAMTools, Picard, the GATK, BEDTools, SnpEff, VCFtools and tabix installations – see **Introduction and Description** above for web links to these free third-party programs). Alternatively, dependencies contained within the `PATH` variable will work automatically with the default settings in `SPANDx.config`. If a "dependency not found" error occurs, please check the installation path and the location specified in the `SPANDx.config` file. Please also specify the location of the PERL5 libraries (automatically installed with VCFtools), which is required for correct functioning of VCFTools, and make sure that the location of tabix and bgzip (dependencies of VCFtools) are specified in your `PATH` variable. By default `SPANDx.config` will automatically alter your `PATH` variable assuming your PERL5LIB and tabix installation paths are `/bin/vcftools_0.1.11/perl` and `/bin/tabix-0.2.6/`, respectively. If you use a different version of this program or have a different install path, please modify. Some of these programs have additional dependencies that are required for them to function properly (e.g. Java). Please refer to the appropriate manuals or system administrator for installation of these utilities if they are not already on your Linux system.

SPANDx customisation:

qsub:

Depending on your cluster environment, the `qsub.config` file may need to be changed. By default, SPANDx will expect the resource manager to be PBS. **IMPORTANT: If you are using SGE, please modify the SCHEDULER variable to SGE**. This file will configure the operation of `qsub` and the variables that SPANDx will run with the `qsub` command line. By default, all `qsub` commands request one node and 12 hours wall time, you will not be sent mail, and standard output is merged with standard error. These settings can be changed if your job needs more time to complete or if you want e-mail notifications of job completion.

Variant calling:

One advantage of SPANDx over other tools is that the GATK variant calling parameters are already specified. These parameters have been tested across NGS data for several bacterial species generated on different NGS platforms. Therefore the default settings should work well for most haploid genome projects. If desired, users can customise the SPANDx variant filtering parameters by altering the `GATK.config` file. All filtering steps used in the GATK can be customised using this configuration file. Note that these variables must conform to JEXL specifications.

SPANDx variant calling has been optimised for bacterial genomes, but may behave differently for other haploid organisms. If in doubt, outputs should be verified with e.g. wet lab analysis of variants. SPANDx is currently not configured to analyse diploid or polyploid genomes but this feature is in development.

The following parameters can be customised to change the variant filtering behaviour for both SNPs and indels if required; below are the default SPANDx settings:

`CLUSTER_SNP=3` (for SNPs only)
`CLUSTER_WINDOW_SNP=10` (for SNPs only)
`MLEAF=0.95`
`QD=10.0`

MQ=30.0
FS=10.0
HAPLO=20.0
QUAL=30.0
LOW_DEPTH=2 (variants with less than the average coverage, divided by LOW_DEPTH, will fail filtering. If this value is set at the default of 2, regions with less than half the average depth will fail and thus will be filtered out).
HIGH_DEPTH=3 (a value of 3 means that regions with more than three times the average coverage of the entire genome will fail and thus will be filtered out).

INTERPRETING THE OUTPUTS

Unaligned reads:

A .bam file of the unaligned reads is generated by SAMTools after BWA alignment. The unaligned reads can be found in: $PWD/strain/unique/unmapped.bam. It is anticipated that future versions of SPANDx will include an option for automated assembly of these unaligned reads.

Alignment files:

Alignment and alignment index files generated with SPANDx are in .bam and .bai format and are found in: $PWD/strain/unique/strain.realigned.ba*. If visualisation of the alignment is desired, these files can be easily viewed in an alignment viewer (our favourite is Tablet [Milne *et al.*, *Bioinformatics* 2010 26(3):401-02; downloadable from: http://ics.hutton.ac.uk/tablet/]).

Whole genome coverage (a.k.a. locus presence/absence):

Following assessment of whole genome coverage by BEDTools, SPANDx provides a combined BEDcov matrix for all analysed genomes in: $PWD/Outputs/Comparative/Bedcov_merge.txt. This file lists the BEDTools windows based on the reference sequence and the corresponding coverage as a percentage of this region for each individual strain, and can be imported into Microsoft Excel for easier visualisation and manipulation. Bedcov_merge.txt is a useful file for identifying the 'core' genome of a given dataset or for identifying variable genetic regions among strains.

Variants:

SNPs and indels are output from SPANDx analysis into two locations: 1) $PWD/Outputs/SNPs_indels_PASS, which contains both SNP and indel variants that have passed the filters specified in the GATK.config file (see "**Variant calling**" above for details of the default filters); and 2) $PWD/Outputs/SNPs_indels_FAIL, which contains SNPs and indels that have failed filtering parameters. If annotation is switched on, annotated variants will be output to: $PWD/strain/unique/annotated. Annotated SNP and indel outputs will be generated for each genome under analysis. In addition, if both annotation and comparative analysis is switched on (-a yes and -m yes), annotated, merged SNP and indel matrices are generated for all genomes under analysis. These files are found in $PWD/Outputs/Comparative and are called All_indels_annotated.txt and All_SNPs_annotated.txt. These are space delimited text files that can be easily imported into Excel. Note that the binary column may need to be specified as “Text only” due to the character string containing “0”. The binary column is a representation of the SNP/indel pattern at that specific location, which can be useful for filtering algorithms.

N.B. SNPs/indels represented with “?” are an ambiguous call and should be interpreted with caution. SNPs/indels represented with “.” do not pass the depth filters and are likely in deleted regions.

SPANDx can repeat the variant filtering steps without repeating the alignment and data processing steps. To use this behaviour, remove the relevant `snps.PASS` and `indels.PASS` files from the `$PWD/Outputs/SNPs_indels_PASS` directory and the relevant `snps.AMB` and `indel.AMB` files from the `$PWD/Outputs/SNPs_indels_FAIL` directory, change the `GATK.config` file to the desired parameters and re-run SPANDx. NB. SPANDx will only re-filter the variants with altered parameters for those strain/s that have been removed from the `Output` directories.

Orthologous SNP matrices for phylogenetic analyses:

Two separate SNP matrix files are generated by SPANDx. These matrices are output in `$PWD/Outputs/Comparative/` and are named `Ortho_SNP_matrix_RAxML.nex` and `Ortho_SNP_matrix.nex`.

> SPANDx excludes SNPs that are low-quality, that are in non-orthologous regions, and that are tri- or tetra-allelic in nature. Non-orthologous SNPs cannot be used for phylogenetic reconstruction, and filtering for tri- and tetra-allelic SNPs is performed to minimise erroneous calls (which can look like tri- and tetra-allelic SNPs) passing through filters.

`Ortho_SNP_matrix.nex` includes SNP coordinates and SNPs identified by the GATK. `Ortho_SNP_matrix.nex` is directly importable into PAUP* and is useful for phylogenetic estimations that require nucleotide data (e.g. maximum likelihood).

`Ortho_SNP_matrix_RAxML.nex` is a RAxML- and PHYLIP-importable version of the `Ortho_SNP_matrix_RAxML.nex` file. Note that for compatibility with PHYLIP, taxa names must have exactly 10 characters (including spaces) otherwise `Ortho_SNP_matrix_RAxML.nex` will not be recognised as a valid PHYLIP file. SPANDx does not automatically rename taxa to meet this PHYLIP requirement.

Log files:

By default, both the standard error and standard output are merged into a single log file. Almost all commands in SPANDx are captured in log files. If an error occurs the log files are a good first place to look.

EXAMPLES

The simplest way to run SPANDx is if your reads are in paired-end, Illumina format and follow the naming convention of `strain_1_sequence.fastq.gz` and `strain_2_sequence.fastq.gz`. SPANDx can then be run by simply specifying the reference `.fasta` genome. All read files in the current directory will be processed, although a SNP or indel matrix will not be constructed nor will variant annotation be performed unless specified.

“No frills” SPANDx command for basic Illumina 1.8+ analysis:

```
/home/user/bin/SPANDx/SPANDx.sh -r reference
```

If other SPANDx features are required or reads other than Illumina v1.8+ are used, these features will need to be specified as per the examples below.

Paired-end Illumina 1.8+ reads, SNP matrix required, no annotated genome available/required:

```
/home/user/bin/SPANDx/SPANDx.sh -r reference -m yes
```

To include an annotation for the above example:

```
/home/user/bin/SPANDx/SPANDx.sh -r reference -a yes -m yes -v
ref_genome_in_SnpEff_database
```

Paired-end Illumina 1.8+ reads, indel matrix required, no annotated genome available/required:

```
/home/user/bin/SPANDx/SPANDx.sh -r reference -i yes –m yes
```

Paired-end Illumina 1.3 reads, SNP and indel matrices required, annotated reference genome Hi_86-028NP.fasta available:

```
/home/user/bin/SPANDx/SPANDx.sh -r Hi_86-028NP -o Hi -m yes –i yes -a
yes -v Haemophilus_influenzae_86_028NP_uid58093 -t Illumina_old
```

Single-end Ion PGM reads, SNP/indel matrices and annotation not required, BEDCov window size of 500bp (instead of the default 1000bp) desired, using the same reference genome as above:

```
/home/user/bin/SPANDx/SPANDx.sh -r Hi_86-028NP -o Hi -t PGM -p SE -w 500
```

Paired-end Illumina 1.3 reads, annotated genome available/required for the reference genome Hi_86-028NP.fasta with a single strain (Hi_00345) for alignment and variant calling. No SNP/indel matrices required:

```
/home/user/bin/SPANDx/SPANDx.sh -r Hi_86-028NP -o Hi -a yes -v
Haemophilus_influenzae_86_028NP_uid58093 -t Illumina_old -s Hi_00345
```

AUTHORS AND CITATION

Dr. Derek Sarovich and Dr. Erin Price, Menzies School of Health Research, Darwin, NT 0810, Australia.

mshr.bioinformatics@gmail.com.

If you find an error or bug please contact the authors on the above e-mail. Please include a detailed description of the error encountered, the operating system used and what happened to cause the error. In addition, please send the appropriate log files with the description.

If you used SPANDx and found it useful for your research, please cite it! ☺

Sarovich DS and Price EP. 2014. SPANDx: a genomics pipeline for comparative analysis of large haploid whole genome re-sequencing datasets. *BMC Res Notes* 7:618.