2D Range expansions

I. Forward simulations

a) Models nomenclature

dimension	# core demes	# of front demes	chr length (Mb)	# length of expansion (demes)	# Founders	migration %	growth time	extra info
2d_	c1_	gr30_	l100_	d100_	f4_	m10_	g5_	mx20_tot_mig

b) Burn in phase

Files

- **params_2d_c5_e0_l100_g0_d0.sh**
  - parameters info to run simulation
- **00_run_model.sh**
  - dispatches model
- **00_burnin_2D_expansion.slim**
  - slim file that defines demographic simulation to run in SLiM
- **00_run_as_slurm_job.sh**
  - sends sim as a SLURM job on cluster (array number == replicate number)

Folder structure

Paste all files into the "mother" folder (where you have differnt models results in different folders) (e.g. `$HOME/2D_fwd_RangeExpansion`)

Dispatch sims

```
sbatch --array=1-100 00_run_as_slurm_job.sh 2d_c5_e0_l100_g0_d0
```

back to top ↑

c) Expansion phase

- **01_run_model_2d.sh**
  - dispatches model
- **01_2D_expansion.slim**
  - slim file that defines demographic simulation to run in SLiM
- **01_slurm_run_model_2d.sh**
  - dispatches as SLURM job
- **00_slurm_set_up_folder_2d_sims.sh**
  - sets up folders (run b4 the sim!!!)
- **param_** files: under folder param_files in the "mother" folder

- ex: **params_2d_c1_gr30_l100_d100_f4_m20_g5_mx20_tot_mig.sh**

Setting up folders

```
script=00_slurm_set_up_folder_2d_sims.sh
model=f4_m20_g5_mx20_tot_mig
sbatch --array=1-200 ${script} ${model}
```

Dispatch sims

`model`: model name `n_samps`: number of individuals to be sampled at generation 3 of growth

```
model=f4_m20_g5_mx20_tot_mig
n_samps=10
script=01_slurm_run_model_2d.sh

dir=/${HOME}/2D_fwd_RangeExpansion/2d_c1_gr30_l100_d100_${model}

cd ${dir}

sbatch --array=1-200 ${script} ${model} ${n_samps}
```

Output/results

VCF files

- `model_folder/replicate_folder`
  - `vcf_files/`
    - " out_gen_25003_p6.vcf.gz"
    - ...
    - " out_gen_25498_p996.vcf.gz"
      - VCF files with the number of sampled individuals determined during dispatch

> - **Note to self**:
>   - Forgot to fix this issue (there is a blank space at the begining of the vcf file names!!!!)

Pop size files

- `model_folder/replicate_folder`
  - `popsizes/`
    - active_demes_`${rep}`.txt.gz
    - active_demes_oneliner`${rep}`.txt.gz
    - popSizes`${rep}`.txt.gz
    - sampled_edges`${rep}`.txt.gz

> this files are just for checking if the demographic model worked properly... **SLiM log file**

- `model_folder/replicate_folder`
  - grid_end_slim_`${rep}`.output
    - log information from SLIM (helps to check migration scheme)

back to top ↑

II. Allele counts

a) Set up

Files

Paste these files inside model folder:

- **02_get_allele_counts_for_pi_calcul.sh**
  - Uses **VCFtools** to obtain allele counts for every VCF file output by the simulations
- **02_slurm_get_allele_counts_for_pi_calcul.sh**
- **sampling_active_edge_demes.txt**
- log folder: **log_allele_counts**

Folder organization

Move all replicate folders to a sub folder named **`sim_files`**: `model_folder/sim_files/replicate_folders`

Input files

- `$model/sim_files/$replicate/out_gen_25498_p996.vcf.gz`
- **sampling_active_edge_demes.txt**: describes which populations and generations were sampled

b) Dispatch & results

On the Cluster

```
model=f4_m10_g5_mx20_tot_mig
dir=${HOME}/2D_fwd_RangeExpansion/2d_c1_gr30_l100_d100_${model}
cd ${dir}
sbatch --array=1-200 02_slurm_get_allele_counts_for_pi_calcul.sh ${model}
cd $HOME
```

Output/results

One file for each sampled generation (total ~ 100 gens, see VCF files output from simulation)
`$model/sim_files/$replicate/vcf_files/count_size_g${gen}_p${pop}_r${rep}.txt.gz`

back to top ↑

III. Calculate Pi (Nucleotide diversity)

a) Set up

Files

- **03_FUNCTIONS_Get_Pi_Profile.R**
  - R function files
- **03_Get_Pi_Profile.R**
  - Main R script
- **03_SLURM_Get_Pi_Profile.sh**
  - Sends analysis as SLURM job
- **03b_SLURM_conCAT_replicate_files.sh**
  - concatenates resulting files (per replicate) into a single file (per model)

Input files

- `$window_size`_win_coord.csv
- `count_size_g${gen}_p${pop}_r${rep}.txt.gz`
- `sampling_active_edge_demes.txt`

b) Dispatch & results

On the cluster

```
model=f4_m10_g5_mx20_tot_mig
dir=${HOME}/2D_fwd_RangeExpansion/2d_c1_gr30_l100_d100_${model}
cd ${dir}
script=03_SLURM_Get_Pi_Profile.sh
founders=4
migration=10
growth=5
samples=10
active=T
old_demes=F
wsize=med
sbatch --array=2-200 ${script} ${founders} ${migration} ${growth} ${samples}
${active} ${old_demes} ${wsize}
echo "4 10 5 10 T F med"
cd $HOME
```

Output files

- `$model/sim_files/$replicate/`
  - `genomic_profile_$wsize/`
    - `genomic_profile_active_demes_f${founders}_m${mig}_g${growth}_r${rep}_resamp_${inds}inds.txt.gzip`

c) Concatenate replicate files

```
model=f4_m10_g5_mx20_tot_mig
dir=/storage/homefs/fs19b061/2D_fwd_RangeExpansion/2d_c1_gr30_l100_d100_${model}
cd ${dir}
founders=4
migration=10
growth=5
n_samps=10
window_size=med
sbatch 03b_SLURM_conCAT_replicate_files.sh ${founders} ${migration} ${growth}
${n_samps} ${window_size}
echo "4 10 5 10 med"
```

Outputfiles

- `$model/`
  - `genomic_profile_$wsize/` - `genomic_profile_all_reps_${model_prefix}_samp_${n_samps}inds_win_${window_size}.txt.gz`

`model_prefix=f4_m10_g5`; → `f${founders}_m${migration}_g${growth}` instead of `model=f4_m10_g5_mx20_tot_mig`;

back to top ↑

IV. Get troughs

a) set up

Files

- **04_slurm_get_troughs.sh**
- **04a_get_troughs.R**
- **04_FUNCTIONS_get_troughs.R**

> **Important NOTE** initial diversity levels are **hard coded** in the R script!!! --> this value changes IF burn-in time and ancestral population size changes. currently: 2500 individuals; 25000 generations of burn-in.

Input files

b) Dispatch & results

On the cluster

```
model_prefix=f4_m10_g5
level=0.1
samps=10
wsize=med

dir=/storage/homefs/fs19b061/2D_fwd_RangeExpansion/2d_c1_gr30_l100_d100_${model_prefix}_mx20_tot_mig
cd ${dir}

sbatch 04_slurm_get_troughs.sh ${level} ${model_prefix} ${samps} ${wsize}
```

Output files

- `summary_data_lvl_${level}_${model_prefix}_${samps}inds_win_${wsize}.txt`
- `complete_data_lvl_${level}_${model_prefix}_${samps}inds_wind_${wsize}.txt`

back to top ↑

V. Graphing results

Use output from previous step (IV) to run graphing scripts... This step is generally done interactivelly in work computer (not the cluster as previously), since depending of the results, graphing params have to be adjusted as needed

a) Genome profiles

Exception: "Genome profile" → can be ran in cluster:

- **04b_slurm_get_genome_profile.sh**
- **NEW_04b_plot_results.R**
- **04_FUNCTIONS_get_troughs.R**

Dispatch

```
script=04b_slurm_get_genome_profile.sh
level=0.1
```

```
model_prefix=f4_m10_g5
samps=10
wsize=med

dir=${HOME}/2D_fwd_RangeExpansion/2d_c1_gr30_l100_d100_${model_prefix}_mx20_tot_mig
cd ${dir}

sbatch ${script} ${level} ${model_prefix} ${samps} ${wsize}
```

back to top ↑

b) Troughs characterization graphs

Files:

- **04b_plot_different_models_together.R**
- **04c_FUNCTIONS_get_troughs.R**
- **04c_plot_diff_models.R**
- **test_functions.R**

back to top ↑