

Status of dead positives recovery

Meeting notes

Gus, Gisella, Andrea

2015-02-04 (Wednesday)

Contents

1	Overview of Discussed	1
2	Current structure of paper	1
3	Gus's proposal to identify LD "outlier" snp-pairs	2
4	Current/future plans	2

1 Overview of Discussed

- Andrea's write up of current results
 - *need to be collated into one doc*
- Status of LD analysis and how to choose cut-off
 - results of discussion with Mark
 - success in generating expected behavior of mean LD behavior vs distance
 - Gus's proposed idea to identify "outlier" snp pairs (*see section below*)
- First figure should have a map of population locations as a panel
- Structure of paper (***mostly*** *unchanged since last meeting*)
- plans for future

2 Current structure of paper

mostly *unchanged since last meeting:*

1. Development of the base SNP set

2. Linkage Analysis

- a. **(new)** linkage-based grouping of contigs by physical proximity

3. Functional Annotation of filtered SNPs

4. Discussion

- establish the ability to do this scale of work in *G. f. fuscipes*
- limits of the dataset as it now stands
- Never-the-less, hypotheses can be formulated and here they are. . .
- **(new)** provides more information pertaining to the physical proximity of the supercontigs

3 Gus's proposal to identify LD "outlier" snp-pairs

1. for each group of SNPs x bp apart: collect r^2 from $\pm \sim 5$ bp distance window around x :
 - a. across genome
 - b. across scaffold
2. calculate modified z-score (based on *median absolute deviation* rather than standard deviation: **MAD is more robust than SD for HTS-type data**)
3. flag any SNP-pair with $z \geq 3.5$
4. possibly randomize data and calculate FDR to evaluate performance.
 - a. perhaps vary the window-size from step 1 to use FDR to chose window-size that minimizes FDR.

4 Current/future plans

Gisella:

- [X] email Washington group about methods of Seq prep and analysis
- [X] email Aksoy group about linkage status of *G. m. morsitans*

Gus and Andrea:

- [] add methods in "final" style to growing document
- [] place document in shared location

Gus:

- [] add LD methods and results to document
- [] generate population-location map with zoom out to all Uganda
- [-waiting-] upon decision of which near-by genes are "interesting" send summary of info known about them to Aksoy group for ideas.