**1. Notes in chronological order**

p. 4: It is true that loss of differentiation can occur if the same allele becomes *fixed* in both divergent populations. But this seems to be much less likely under divergent selection than loss of differentiation due to loss of the locally beneficial de-novo mutation. At least, I would formulate this in terms of extinction, rather than fixation. A related question would be how long it takes for neutral divergence to decay after fixation of the same allele at the locus temporarilly under divergent selection. [I consider these very minor comments]

p. 4: Aeschbacher & Bürger (2014) seems to have been misinterpreted. We did not compute the probability of loss (or rate of stochastic loss), but the expected survival time of a de-novo mutation. They are related, but not the same. [Minor remark]

p. 5: Apparent conflict between Feder et al. (2012) and Yeaman and Whitlock (2011):

* **Mutation models.** Feder et al. (2012): Within a simulation experiment, the selection coefficients at the background loci (s\_0) and of the *de novo* mutation (s\_n) were each kept at a given value. Yeaman and Whitlock (2011): mutation effect sizes are drawn from a Gaussian distribution. Effects of a new mutation are added to the effect of the previous allele. There is recurrent mutation at a constant per-locus rate. Under this continuum-of-alleles mutation model, a single locus can effectively represent a single locus with recurrent mutation, or multiple, completely linked, individual loci.
* **Divergent directional selection vs. divergent stabilising selection** (i.e. difference in the fitness regime). Feder et al. (2012): divergent *directional* selection with no dominance. Yeaman and Whitlock (2011): *stabilising* selection for diverged optima; dominance and epistasis occur as a consequence of the curvature of the fitness function (gamma). Simulations do account for the stochastic loss of mutations; in fact, given the model of selection, loss of mutations is necessary to explain the results found by Yeaman and Whitlock. Also, at the end of subsction “Linkage between loci”, Yeaman and Whitlock (2011) say that “As such, although we expect strong selection to maintain loosely clustered architectures, beyond a certain point there will be little fitness benefit of further increasing linkage”. In other words, their prediction is consistent with a diminishing return of linkage as adaptive divergence increases, or as speciation proceeds.
* **The input DFE** may play an important role, but is so far underexplored. Note that Yeaman and Whitlock (2011) highlight that mutation effect size and mutation rate are expected to strongly influence the dynamics of establishment of new mutations.
* Yeaman and Whitlock (2011) disinguished between “transient” and “stable” to differentiate between short- and long-lived migration–selection polymorphisms, respectively. Their distinction criterion was arbitrary, but it illustrates that the simulations they performed did indeed feature the **“stochastic loss”** aspect that is being promoted by Rafajlovic et al. (2016).
* Fig. 3A for the **‘low-migration’ scheme** shows that the system cycles over different allelic configurations that replace each other, but at the phenotypic level lead to a constant level of divergence. This process includes both the establishment as well as stochastic loss of locally adaptive mutations. At later stages, the genetic architecture does become more concentrated, with individual clusters more resistant to stochastic loss. Overall, this **illustrates nothing but the competition of genetic architectures** suggested by Yeman and Whitlock (2011)
* Yeaman and Whitlock (2011) emphasise that **mutation rate has a strong effect** on the stability of concentrated architectures at larger population sizes (i.e. when drift is less important).
* In their Figure 7, Yeaman and Whitlock (2011) show that in small populations and with weak selection, there is no concentration of the genetic architecture, as genetic drift and migration prevent the long-term maintenance of divergence. That is, in this condition, the stochastic loss argument predominates the process of establishment. The importance of **stochastic loss in the sense of drift dominating selection has therefore already been recognised** by Yeaman and Whitlock (2011).
* In their Discussion (p. 1907–1908), Yeaman and Whitlock (2011) explicitly mention the **“homogenisation of previously established divergent alleles at more loosely linked loci”** as a way of how concentrated architectures can arise, i.e. how diffuse architectures are replaced by more concentrated architectures.
* **Yeaman (2013) does not “artificially” exclude the possibility of the emergence of clusters** of differentiation.
* The main reason for why Feder et al. (2012) conclude that clustering is less important as divergence proceeds, while Yeaman and Whitlock (2011) see the evolution of concentrated architectures is twofold:
  + **different models of selection** (directional vs. stabilising different optima)
  + different mutation models (single mutation vs. continuum-of-mutations model)

p. 6

* The sentence “Critically, the existing theory relies on the assumption that any established local genomic divergence persists indefinitely, or increases due to the accumulation of new beneficial mutations.” simply is not true. The **simulations by Yeaman and Whitlock (2011) readily account for the loss of mutations**, and Aeschbacher and Bürger have approximated the mean extinction time for locally beneficial de-novo mutations. The **authors contradict themeselves** when they say that Aeschbacher and Bürger (2014) considered the loss of locally beneficial mutations, while later they say that previous literature only focussed on establishment and accumulation of divergence.
* The authors imply that **Yeaman and Whitlock (2011)** did not find evidence for clustering of differentiated loci early on under weak selection and high migration. But this is exactly what their Figures 3B,D and 5C show: under weak selection and with tight linkage, the clustering distance (an inverse measure of the extent of clustering) decreases with increasing migration. **Yeaman and Whitlock (2011) did not need to invoke any mechanism that reduces recombination** to explain this.
* The observation of the authors that clusters grow rapidly during early stages of divergence, but shrink in size during the late stage, is nothing but compatible with Yeaman and Whitlock (2011), as it suggests **the evolution of *increased* clustering in terms of more narrow clusters**.
* Yeaman and Whitlock's (2011) results are also in agreement with the authors’ obervation that, under strong selection, clustering is less abundant early on, but may then happen over longer time scales. **If selection is strong, individual locally beneficial mutations are, on average, less dependent on being linked to a background locus**. There may still be a small benefit to be gained by increased linkage, but it takes more time for this to translate into concentration.

p. 10

* Are the authors **equating the width of a cluster with the strength of clustering?** That would not be appropriate!

p. 11

* The authors repeatedly suggest that a highly polygenic architecture of the selected trait automatically implies that there is little establishment bias, and a not-so polygenic trait implies a high establishment bias. But there are many other factors: strength of selection, distribution of fitness effects, migration rate, model of selection that are just as decisive.

p. 13

* The authors state that “when the rate of loss at a locus is larger than the rate of gain at this locus in a tiven stage of divergence, this locus is unlikely to make a sustained contribution to overall divergence”. This is a **narrow view of “contribution to divergence”**, as a given locus may well contribute a transient local reduction in effective gene flow by which other mutations can more easily establish, even though they may later be lost again because the architecture has become more concentrated around another diverged locus.

p. 19

* In the two-locus gain-loss model, the size of a cluster is defined as twice the recombinational distance at which the rate of establishment becomes smaller than the rate of loss at the weaker locus.
* The observation in both the two- and the mulitlocus model that the cluster first increases and then shinks is i) consistent with the idea of a competition of architectures resulting in increasing concentration; ii) might be amplified here due to the model of selection **(stabilising selection towards** **different optima). This model of selection imposes a regime** of diminishing returns.
* The sentence “Reproductive isolation between populations is most efficient when many small barriers to gene flow are formed throughout the genome” seems completely off. Reproductive isolation can be very strong at a single locus, with the rest of the genome exchanging material. It seems as if reproductive isolation is being mixed up with genomic divergence.

p. 20

* The first paragraph reads weird, given that previous literature has already suggested that weak to moderate selection per locus and strong migration are most favourable to the evolution of clusters. Again, the **works by Yeaman and Whitlock (2011), and Feder et al. (2013) and Yeaman (2013) are wrongly interpreted**.

**2. Todos / open questions**

* Approximate the stabilising fitness function by a linear function (for directional selection), assuming that the indivdidual is far from the fitness optimum.
* Obtain the probability of loss of an established mutation, accounting for a genomic background that consists of more than one locus. Isn't it the case that, with an increasing number of migration–selection polymorphisms (MSPs) in the background, it becomes less likely that an established mutation goes extinct? If so, this would limit the importance of stochastic loss, and predict that stochastic loss imposes less of a limit as speciation progresses. It would therefore also predict that clustering becomes less important as speciation progresses.
* What would be a good metric to measure the extent of clustering?
* Could the way genomic divergence is measured (i.e. in terms the effect-size difference *D* between the most frequent alleles) be problematic? With a population of size 1000 and a mutation rate of 2\*10^(-5) per locus and individual (as often used in Rafajlovic et al. 2016), there would on average be 2\*10^(–2)\*10^5 = about 300 mutations at a given site over 10^5 generations. This will result in some amount of stacking, and to competition among multiple alleles at the very same site for establishment. Not sure if this affects *D* much.
* The authors define “successful establishment of a mutation” to mean that the mutation is most common (frequency > 50 %) in the deme where it is advantageous. While this definition seems fine in the context of biallelic loci, it is problematic here with potentially more than two mutations segregating at the same locus. The deterministic equilibrium frequency of a locally beneficial mutation may well be considerably lower than 0.5, but the mutation may nevertheless have established in the sense that it has survived the *early* phase of stochastic loss. Therefore, I wonder if the Rafajlovic et al. (2016) might be substantially underestimating the establishment probability. [This might be an important issue]

**3. Scope of a potential paper**

1. Address the various misconceptions. Provide a verbal explanation of the underlying processes and clarification.
2. Set up an efficient, approximate simulation scheme to model the dynamics of establishment and extinction in a multi-locus context. This will allow to jointly study the “establishment bias” and “stochastic loss” in a multi-locus context and for a large parameter space. Challenge: how to compute the equilibrium allele frequencies at a given MSP? Can we do without, e.g. using the effective migration rate at that locus if it were neutral, and then model it as if it was the only locus under selection? We may want to do this in the continent–island model.
3. Introduce an appropriate metric to measure the extent of clustering. The goal is twofold. First, provide a means of quantifying the amount of clustering for empiricists. Second, quantify the amount of clustering in excess to the one expected without linkage. Assume that each MSP is established without the synergistic help of others. This is the null model against which one can judge the level of clustering induced by linkage. Throughout, assume a constant per-base pair mutation rate. The prediction under this null model is independent of the distribution of selection coefficient, as long as this distriubtion is constant across time.
4. Repeat some of the analyses by Rafajlovic et al. (2016) using a more appropriate criterium for establishment than that the alleles has reached a frequency of at least 0.5.