

Predicting Phenotypic Output Given Genotype and Assessing The Implications on Susceptibility to Infection

Using Negative Binomial Distribution Models to Analyze Count Data on Legume Nodules

Table of Contents

Abstract	2
Background	3
<i>Inspiration</i>	3
<i>Biological Context</i>	3
<i>Description on Data</i>	4
Models and Methodology (All Models in Appendix)	6
<i>HM277 NBD Model</i>	6
<i>HM310 Model</i>	7
<i>NBD Model Across All Genotypes</i>	8
<i>NBD Model Across All Genotypes, Zero-Inflated</i>	9
Conclusion	10
<i>Parameter Summary</i>	10
<i>Key Takeaways, Limitations, and Next Steps</i>	11
Appendix A:	13
Appendix B: HM277 NBD Model	14
Appendix C: HM310 NBD Model	15
Appendix D: All Genotypes NBD Model	16
Appendix E: All Genotypes NBD Model, Zero-Inflated	17
Appendix F: LRT Test Between Spiked and Unspiked Model	18

Abstract

All organisms have a unique genetic code made from a combination of alleles, representing their genotype. Genotypes are inherited from the parents and determine phenotypes, the observable traits and characteristics of an individual. This analysis utilizes a dataset collected by Dr. Corlett Wood¹, Assistant Professor of Biology here at the University of Pennsylvania, that documented the count of nodules and galls on various legume and nematode genotypes. In the study, the dataset was used to explore a tradeoff between mutualistic and parasitic relationships in the legumes microbiome from an evolutionary standpoint. For my purposes, I use it as a count dataset to apply an NBD model on, where each plant has a different count of nodules (phenotype) and a specific genotype.

First, I run NBD models on the two most different genotypes, HM277 and HM310, to see whether their phenotypes can be predicted given their genotypes, as well as to compare their individual heterogeneities. Second, I apply the NBD model across the entire dataset for all genotypes to begin comparing the individual heterogeneities with the group heterogeneity. Finally I run a zero-inflated NBD model to capture the cases where the legumes failed to form a mutualistic relationship with the nodule-creating bacteria (rhizobia), allowing me to better compare the genotype heterogeneities with the heterogeneity of the aggregate dataset.

Overall, I discovered that the probability distribution of phenotypic outcomes can be reasonably predicted given a genotype. I also was able to quantifiably show the increase in heterogeneity when taking the aggregate of many genotypes and observe the positive effects that such genetic variation had on plant growth.

¹ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6121810/>

Background

Inspiration

Count processes are everywhere, including science. I was compelled to test the robustness of the NBD model in a domain where it may not typically be used, and what better place to source that data than the biology department right here at Penn. Although the biological context is dense with this one, I found it a rewarding journey to dive into Dr. Corlett Wood's research and understand the beautiful coevolved dynamic between legumes, nitrogen-fixing rhizobia, and parasitic nematodes.

There was a fascinating feedback loop between the biology and the statistical model. The biological concepts informed the initial model, but then mismatches in the model flagged holes in my biological understanding, and then by grasping specific biological nuances led to updates in the statistical model and improvements to its performance, and so on and so forth. Overall, I aim to follow a similar structure to this analysis, describing the biological context first and the resulting model, and then updating the model by uncovering nuances in the biology.

Biological Context

Let's go through some definitions:

- **Legumes** are what we are studying; they are one of the few plants that can produce nitrogen in-house by developing mutualistic relationships with nitrogen-fixing bacteria. This experiment uses legumes of the species *Medicago truncatula*. Within the species, there are infinite amounts of unique genotypes that affect the potential for **nodule** growth, which is the observable phenotype.

- **Nodules** are structures where nitrogen fixation occurs. This is the process where atmospheric nitrogen is converted into nitrates and nitrites, which are usable by plants. This process is done by nitrogen-fixing bacteria of the genus **rhizobia**. These positive relationships with bacteria are similar to the positive bacteria that live in our stomach, they are called mutualistic relationships.
- Thus, genotypes that affect the potential for nodule growth are in essence increasing the **susceptibility to infection** from the beneficial rhizobia bacteria.
- However, the same increase in susceptibility to infection from mutualistic organisms also increases the susceptibility to infection from parasitic organisms, like **nematodes**.
- This means that the same genotype that increases the potential for nodule growth (good) also increases the potential for nematode infection (bad), and that extreme nodule growth is a sign of high infection susceptibility and low nodule growth will inhibit ability to grow.
- **Biomass** is the amount of mass above the ground for the legume plant, and measures how well it was able to grow (nodule growth is good for biomass in the sense it increases nitrogen fixation but it is bad for biomass in the sense it increases susceptibility for infection).

Description on Data

The data was collected by a study from the Wood Lab of Evolutionary Biology at The University of Pennsylvania. It was published in a paper called “Genetic conflict with a parasitic nematode disrupts the legume–rhizobia mutualism” on June 2018 by Corlett W. Wood, Bonnie L. Pilkington, Priya Vaidya, Caroline Biel, and John R. Stinchcombe. The methodology is described

on page 235, legume plants of various genotypes were inoculated by rhizobia bacteria in order to create the mutualistic relationship that would allow nodules to grow. They were then placed in soil infected by nematodes to see the interaction between the parasitic relationship and the mutualistic one. The dataset is too large to place here, but is accessible from the footnote on page 2. The following is a summary of the dataset:

Sample Size	388
Genotypes	10
Average # of Nodules	19.94
Average Biomass	0.13 grams

Figure 1; sourced from dataset

The relevant features include what genotype the plant is, the amount of nodules the plant has, and the biomass of the plant in grams which measures how much the plant grew in total. Additionally, the following is a breakdown of average nodules and biomass by genotype:

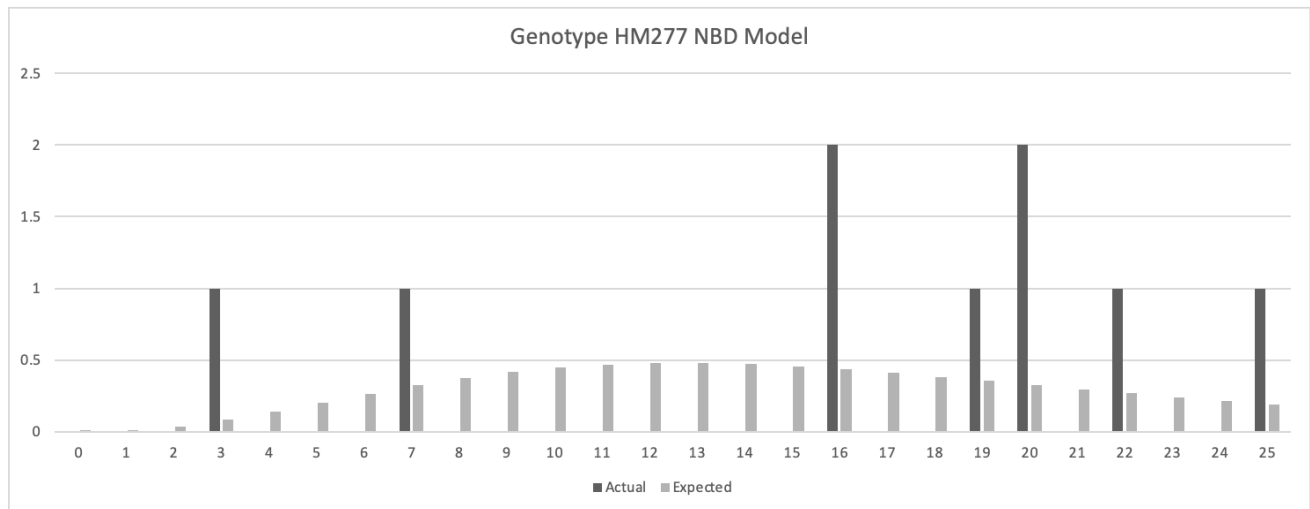
Genotype	Average # of Nodules	Average Biomass (g)
HM259	15.93	0.13
HM260	19.19	0.14
HM270	18.73	0.12
HM271	21.37	0.13
HM276	21.38	0.13
HM277	14.53	0.11
HM287	21.34	0.12
HM302	19.33	0.13
HM307	19.14	0.13
HM310	21.72	0.11

Figure 2; sourced from dataset

Models and Methodology (All Models in Appendix)

Since I am trying to model the count of nodules for specific genotypes and see the difference in distributions, it seems reasonable to take the most extreme cases. This will also enable me to draw better conclusions. From Figure 2, HM277 has the lowest nodule count and HM310 has the highest nodule count. Thus, I will take the NBD models of these two genotypes.

HM277 NBD Model



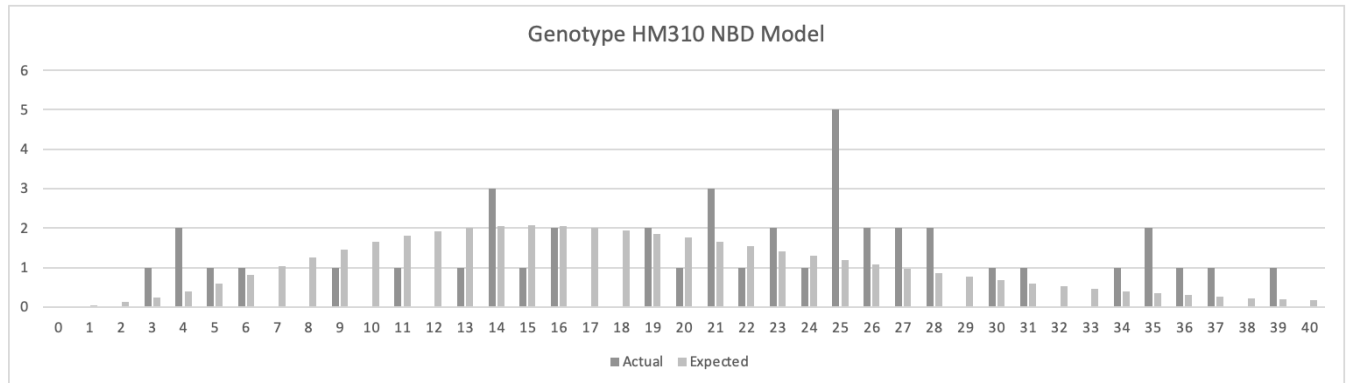
*Note there were no plants of genotype HM277 that had a nodule count higher than 25, so I decided to cut the model off there

r	4.489	chi-sq	38.268
alpha	0.291	df	26
Total LL	-31.32	p-value	0.024

On an initial look, the histogram does not appear to fit the data well. Additionally, we are met with a high chi-squared and low p-value despite the degrees of freedom being relatively high. Still, a p-value of 0.024 is not something to overlook, and the upcoming models reveal there is still value in an NBD approach.

The biggest takeaway can be drawn from the high r value of 4.489. This aligns with the understanding that homogeneity would be far higher in a specific genotype. It suggests that plants of the same genotype are spinning roughly the same wheel.

HM310 Model



*Note there were no plants of genotype HM310 that had a nodule count higher than 40, so I decided to cut the model off there

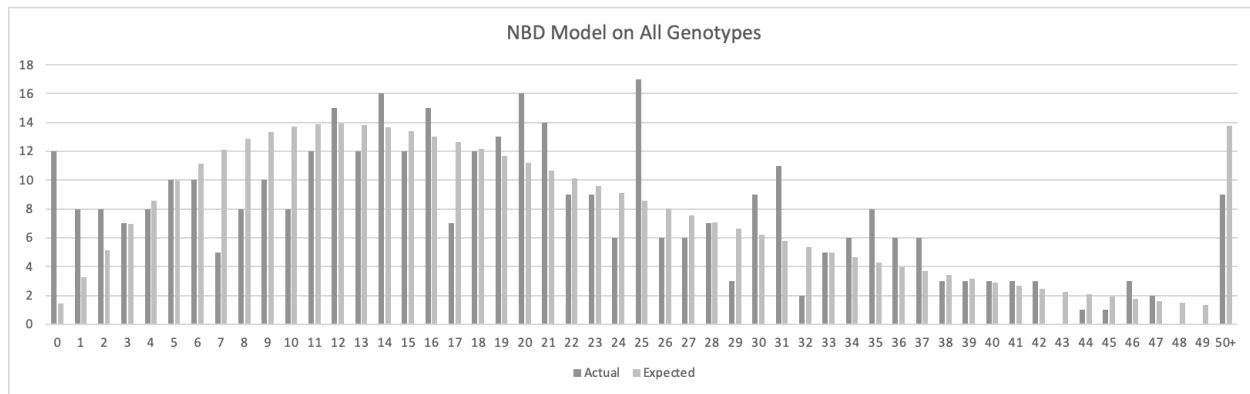
r	5.695	chi-sq	39.805
alpha	0.304	df	38
Total LL	-128.477	p-value	0.390

For this genotype, the histogram seems to track the actual data better. There seems to be room for improvement, but the p -value is much higher at 0.390 which suggests our model is a good fit. Also, it is important to note that drawing comparisons between the two genotype models with regards to LL and χ^2 is a poor idea because they are dependent on the number of samples in each genotype. With that in mind, an explanation for the low p -value in the HM277 model can be the low sample size of that genotype - there were 43 HM310 plants and only 9 HM277 plants. Another note of caution is that because HM310 had much higher nodule counts,

it had more count categories to go into the model, thus increasing its degrees of freedom and its p-value. Finally from its high r value of 5.695, we can see that HM310 is even more homogenous than HM277.

NBD Model Across All Genotypes

The previous models have shown the NBD is an adequate representation of nodule counts. Next, we want to see the change in heterogeneity for the entire dataset when we incorporate all the genotypes.



*Note there were a few plants with nodule counts much higher than 50, so I included a 50+ bucket to incorporate them

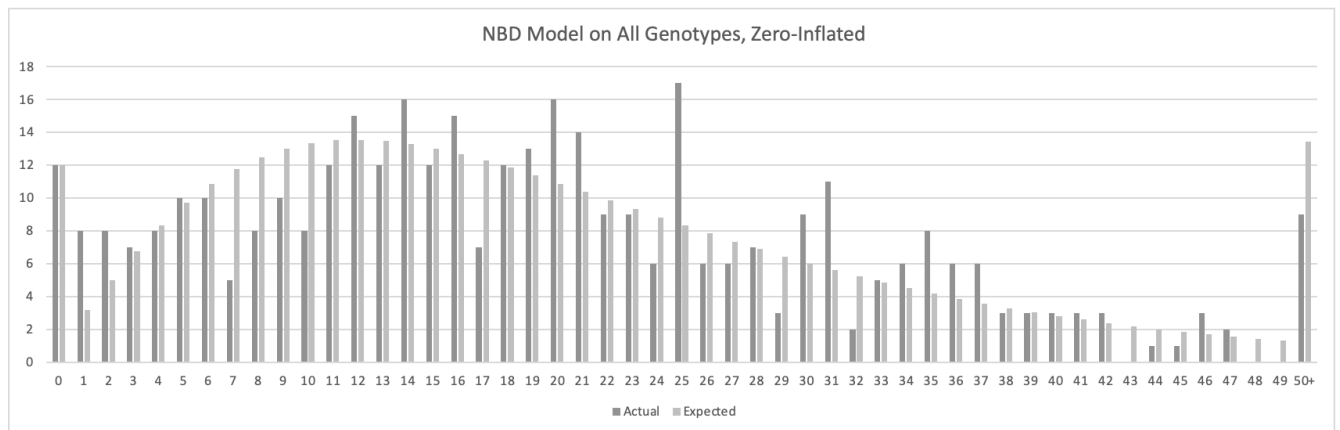
r	2.033	chi-sq	92.558
alpha	0.103	df	48
Total LL	-1474.983	p-value	0.00011

At this point in the analysis, I was concerned. The initial two models on specific genotypes showed that the NBD model had promise in predicting module counts. Now, for a model with significantly more data and fundamentally the same behavior, the p-value was almost

0. The reason for this stems from the relationship between legumes and their mutualistic partner the rhizobia.

Recall from the biological context that legumes can only grow nodules when infected by rhizobia. If the rhizobia fails to infect the plant, then there will be zero nodule growth. It turns out that 3% of the plants in the dataset were not infected by rhizobia and thus were not able to grow nodules in the first place. Regardless of what genotype these plants were, their nodule growth was going to stay at 0, leading to the significantly higher ‘Actual’ number of zeroes than ‘Expected’ number of zeroes. I simply got lucky in the previous two models that genotypes HM277 and HM310 did not have any of these non-infected legumes, but for the aggregate model, we must add a spike at 0 to account for them.

NBD Model Across All Genotypes, Zero-Inflated



r	2.558	chi-sq	61.960
alpha	0.127	df	48
Spike @ 0	0.028	p-value	0.085
Total LL	-1464.218		

Now, the model histogram and the p-value are in a much better range. The spike parameter of 0.028 estimates that 2.8% of the plants were in the ‘non-infected’ category, and the data supports that. Furthermore, the likelihood ratio test provides a low p-value of 3.481E-06 which suggests the model did make a meaningful improvement after spiking at 0. Finally, from the r value, we can see that the aggregate data set of all the genotypes is still fairly homogenous, although far less homogenous than a singular genotype.

Conclusion

Parameter Summary

	HM277	HM310	All Genotypes	All Genotypes Zero-Inflated
r	4.489	5.695	2.033	2.558
Alpha	0.291	0.304	0.103	0.127
Spike @ 0	-	-	-	0.028
Total LL	-31.32	-128.477	-1474.983	-1464.218
Chi-sq	38.268	39.805	92.558	61.960
df	26	38	48	48
p-val	0.024	0.390	0.00011	0.085

Key Takeaways, Limitations, and Next Steps

Genotype	Average # of Nodules	Average Biomass (g)
HM259	15.93	0.13
HM260	19.19	0.14
HM270	18.73	0.12
HM271	21.37	0.13
HM276	21.38	0.13
HM277	14.53	0.11
HM287	21.34	0.12
HM302	19.33	0.13
HM307	19.14	0.13
HM310	21.72	0.11

*Including figure 2 again for easy reference

Our results confirm that genetic variation, which can be measured by the heterogeneity of the nodule count distribution, leads to higher plant growth. Recall that nodule counts are not necessarily good or bad, while a high nodule count does lead to more nitrogen available to the plant, it also increases susceptibility to infection by nematodes. This can be verified from Figure 2, where the extreme ends of the nodule counts, being genotype HM277 and HM310 also have the lowest average biomass.

In the models that aggregated all the genotypes, the r value is significantly lower indicating more heterogeneity across the plants. This makes sense given that having multiple genotypes together means that everyone is spinning different wheels. In theory, this should lead to a more optimal number of nodule counts that balances the risk of infection with the need to fix nitrogen. Checking with Figure 1, the average biomass across the population is 0.13 which is higher than the genotype with the most nodules (HM310) and higher than the genotype with the least nodules (HM277). Thus, we can confirm that the higher heterogeneity in a population of legumes, the higher their average biomass.

The spike at 0 also has an interesting interpretation. It indicates that roughly 2.8% of a population of legumes are ‘non-infectable’ meaning they were not able to develop a colony of mutualistic rhizobia bacteria. This prevents them from being able to fix nitrogen, and thus leads them to having much lower biomasses.

Finally, we can extend the takeaways beyond the world of legumes, rhizobia, and nematodes. The NBD count models were sufficient in predicting the phenotypic characteristics like nodule counts on plants given their genotype. While the p-values were not always high, given enough degrees of freedom and necessary adjustments to the model such as inflating for individuals that can’t exhibit a certain phenotype, the models proved capable of predicting the phenotype count. This suggests that for countable phenotypes of any species, reasonable models can be created using the NBD. Next steps may include applying NBD models on to other countable phenotypes and seeing if similar patterns hold true.

There were two main limitations. Certain genotypes such as HM277 did not have enough data to provide a robust model which may have contributed to a low p-value. Second, I was unsure what to do in situations where genotypes had no counts beyond a certain number. I decided to cut the models off at the number where they had no more counts (or create a 50+ bin if there were many counts greater than 50). This was to prevent the p-value from getting artificially inflated by high degrees of freedom when many of those counts were 0. The tradeoff is that it became more difficult to compare genotype distributions.

Appendix A:

All raw data can be found in the zip folder of the 'Supplementary Materials' section of this link

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6121810/>

Appendix B: HM277 NBD Model

r	4.48782685								
alpha	0.29063281								
Spike @ 0	1E-08								
				-31.315818					
x	n_x	P(X=x)	w Spike	LL	Expected	Chisq		chisq	38.2569431
0	0	0.001242569	0.00124258	0	0.01118321	0.01118321		df	23
1	0	0.004320696	0.00124257	0	0.01118312	0.01118312		pval	0.023900928
2	0	0.009185894	0.0043207	0	0.03888627	0.03888627			
3	1	0.01539206	0.00918589	-4.6900862	0.08267305	10.1785134			
4	0	0.022324916	0.01539206	0	0.13852854	0.13852854			
5	0	0.029363894	0.02232492	0	0.20092425	0.20092425			
6	0	0.035977124	0.02936389	0	0.26427504	0.26427504			
7	1	0.041764877	0.03597712	-3.324872	0.32379411	1.4121764			
8	0	0.046468259	0.04176488	0	0.37588389	0.37588389			
9	0	0.049957196	0.04646826	0	0.41821432	0.41821432			
10	0	0.052208033	0.0499572	0	0.44961476	0.44961476			
11	0	0.053277666	0.05220803	0	0.46987229	0.46987229			
12	0	0.053278468	0.05327767	0	0.47949899	0.47949899			
13	0	0.052356317	0.05327847	0	0.47950621	0.47950621			
14	0	0.050672718	0.05235632	0	0.47120685	0.47120685			
15	0	0.048391167	0.05067272	0	0.45605445	0.45605445			
16	2	0.045667457	0.04839117	-6.056876	0.4355205	5.61993314			
17	0	0.042643328	0.04566746	0	0.41100711	0.41100711			
18	0	0.039442859	0.04264333	0	0.38378995	0.38378995			
19	1	0.036170921	0.03944286	-3.2329023	0.35498572	1.17200041			
20	2	0.032913169	0.03617092	-6.6389996	0.32553829	8.61287948			
21	0	0.029737055	0.03291317	0	0.29621852	0.29621852			
22	1	0.026693491	0.02973705	-3.5153614	0.26763349	2.00408664			
23	0	0.023818871	0.02669349	0	0.24024142	0.24024142			
24	0	0.021137208	0.02381887	0	0.21436983	0.21436983			
25	1	0.01866226	0.02113721	-3.8567204	0.19023487	3.4468946			
	9								

Appendix C: HM310 NBD Model

r	5.6952858								
alpha	0.3041991								
Spike @ 0	0.00001								
				-128.4762					
x	n_x	P(X=x)	w spike	LL	Expected	Chisq		chisq	39.8050649
0	0	0.00025091	0.0002509	0	0.0107891	0.0107891	df		38
1	0	0.0010957	0.0010957	0	0.0471147	0.0471147	pval		0.38968721
2	0	0.00281247	0.0028124	0	0.1209348	0.1209348			
3	1	0.00553155	0.0055315	-5.197297	0.2378543	2.4421085			
4	2	0.00921991	0.0092198	-9.372799	0.3964523	6.4859373			
5	1	0.01370798	0.0137078	-4.289787	0.5894374	0.2859704			
6	1	0.01873574	0.0187356	-3.977332	0.8056287	0.0468953			
7	0	0.02400157	0.0240013	0	1.0320573	1.0320573			
8	0	0.0292044	0.0292041	0	1.2557766	1.2557766			
9	1	0.03407477	0.0340744	-3.379208	1.4652005	0.147701			
10	0	0.03839433	0.0383939	0	1.6509396	1.6509396			
11	1	0.04200488	0.0420045	-3.16998	1.8061916	0.3598427			
12	0	0.04480933	0.0448089	0	1.9267818	1.9267818			
13	1	0.04676692	0.0467664	-3.062589	2.0109573	0.5082329			
14	3	0.04788493	0.0478845	-9.116893	2.0590316	0.4300185			
15	1	0.04820877	0.0482083	-3.032224	2.0729562	0.555359			
16	2	0.04781163	0.0478112	-6.080993	2.0558795	0.0015188			
17	0	0.04678494	0.0467845	0	2.0117323	2.0117323			
18	0	0.04522987	0.0452294	0	1.9448649	1.9448649			
19	2	0.04325036	0.0432499	-6.281519	1.859747	0.0105772			
20	1	0.04094774	0.0409473	-3.195469	1.7607351	0.3286797			
21	3	0.03841671	0.0384163	-9.777818	1.651902	1.1001672			
22	1	0.03574276	0.0357424	-3.331417	1.5369235	0.187574			
23	2	0.03300063	0.0330003	-6.822477	1.4190129	0.2378738			
24	1	0.03025364	0.0302533	-3.498149	1.3008935	0.0695959			
25	5	0.02755378	0.0275535	-17.95813	1.1848008	12.285394			
26	2	0.02494225	0.024942	-7.382404	1.0725061	0.8020886			
27	2	0.02245033	0.0224501	-7.59292	0.9653544	1.1089105			
28	2	0.02010045	0.0201003	-7.814046	0.8643108	1.4922755			
29	0	0.01790742	0.0179072	0	0.7700115	0.7700115			
30	1	0.01587956	0.0158794	-4.142733	0.6828141	0.1473415			
31	1	0.01401986	0.0140197	-4.26729	0.6028479	0.2616411			
32	0	0.01232707	0.012327	0	0.5300589	0.5300589			
33	0	0.01079665	0.0107965	0	0.4642515	0.4642515			
34	1	0.0094216	0.0094215	-4.664761	0.4051245	0.8735013			
35	2	0.00819316	0.0081931	-9.608931	0.3523023	7.7061864			
36	1	0.00710148	0.0071014	-4.947462	0.3053608	1.5801758			
37	1	0.00613607	0.006136	-5.093581	0.2638484	2.0539029			
38	0	0.00528619	0.0052861	0	0.2273039	0.2273039			
39	1	0.00454118	0.0045411	-5.394578	0.1952689	3.3164127			
40	0	0.00389069	0.0038907	0	0.167298	0.167298			
	43								

Appendix D: All Genotypes NBD Model

r	2.03312612								
alpha	0.10333414								
Spike @ 0	0								
				-1474.98337					
x	n_x	P(X=x)	w spike	LL	Expected	Chisq			
0	12	0.0081097	0.0081097	-57.7763402	3.1222326	25.2430758	chisq	92.55801328	
1	8	0.01494383	0.01494383	-33.6276566	5.75337282	0.87728257	df	48	
2	8	0.0205407	0.0205407	-31.0827766	7.90816888	0.00106636	pval	0.000118934	
3	7	0.02502815	0.02502815	-25.8142794	9.63583662	0.72102039			
4	8	0.02854299	0.02854299	-28.4507513	10.9890511	0.81302986			
5	10	0.03121511	0.03121511	-34.6685315	12.0178157	0.33879535			
6	10	0.03316308	0.03316308	-34.0631795	12.7677873	0.59999797			
7	5	0.03449327	0.03449327	-16.8349549	13.2799099	5.16245279			
8	8	0.03530006	0.03530006	-26.7509652	13.5905227	2.29968668			
9	10	0.03566664	0.03566664	-33.3353962	13.7316545	1.01409813			
10	8	0.03566594	0.03566594	-26.668473	13.7313866	2.39224146			
11	12	0.03536164	0.03536164	-40.1055321	13.6142312	0.19139843			
12	15	0.03480909	0.03480909	-50.3681501	13.4015003	0.19066533			
13	12	0.03405624	0.03405624	-40.5569037	13.1116528	0.0942499			
14	16	0.03314446	0.03314446	-54.5100761	12.7606163	0.82234325			
15	12	0.0321093	0.0321093	-41.2633155	12.3620799	0.01060516			
16	15	0.03098119	0.03098119	-52.115624	11.9277594	0.79131899			
17	7	0.02978607	0.02978607	-24.5960015	11.4676363	1.74053082			
18	12	0.0285459	0.0285459	-42.6749033	10.9901717	0.09278774			
19	13	0.02727921	0.02727921	-46.8211944	10.5024965	0.59390865			
20	16	0.02600151	0.02600151	-58.3936098	10.0105819	3.58352091			
21	14	0.02472569	0.02472569	-51.7987768	9.5193894	2.10894528			
22	9	0.02346235	0.02346235	-33.7712244	9.03300482	0.00012059			
23	9	0.02222014	0.02222014	-34.2608038	8.55475579	0.02317336			
24	6	0.02100602	0.02100602	-23.1776786	8.08731589	0.53873098			
25	17	0.01982544	0.01982544	-66.6534144	7.63279618	11.4957226			
26	6	0.01868266	0.01868266	-23.8809555	7.19282564	0.19781281			
27	6	0.01758084	0.01758084	-24.2456751	6.76862174	0.08728208			
28	7	0.01652221	0.01652221	-28.7213467	6.36105219	0.06418031			
29	3	0.01550828	0.01550828	-12.499143	5.97068894	1.47805268			
30	9	0.01453988	0.01453988	-38.0777383	5.59785522	2.06768282			
31	11	0.01361732	0.01361732	-47.2605439	5.24266656	6.32252461			
32	2	0.01274043	0.01274043	-8.72594946	4.90506627	1.72054964			
33	5	0.01190872	0.01190872	-22.1524229	4.58485629	0.0375899			
34	6	0.01112136	0.01112136	-26.9933259	4.28172379	0.68955246			
35	8	0.01037731	0.01037731	-36.5450693	3.99526407	4.01423024			
36	6	0.00967533	0.00967533	-27.8290585	3.72500027	1.38942911			
37	6	0.00901403	0.00901403	-28.2538406	3.47040014	1.8438437			
38	3	0.00839192	0.00839192	-14.3414566	3.23089047	0.01650022			
39	3	0.00780745	0.00780745	-14.5580297	3.0058692	1.146E-05			
40	3	0.007259	0.007259	-14.7765388	2.79471571	0.01507904			
41	3	0.00674493	0.00674493	-14.9968908	2.59679948	0.06260424			
42	3	0.0062636	0.0062636	-15.218999	2.41148723	0.14362393			
43	0	0.00581337	0.00581337	0	2.2381489	2.2381489			
44	1	0.00539263	0.00539263	-5.2227221	2.07616249	0.55782036			
45	1	0.00499979	0.00499979	-5.29835995	1.92491803	0.44442067			
46	3	0.0046333	0.0046333	-16.1234574	1.7838207	0.82917084			
47	2	0.00429167	0.00429167	-10.9021583	1.65229329	0.073171			
48	0	0.00397345	0.00397345	0	1.52977806	1.52977806			
49	0	0.00367724	0.00367724	0	1.41573814	1.41573814			
50+	9	0.04347956	0.04347956	-28.2191793	16.7396302	3.57844675			
	385								

Appendix E: All Genotypes NBD Model, Zero-Inflated

r	2.55833443								
alpha	0.12668487								
Spike @ 0	0.02753762								
				-1464.21778					
x	n_x	P(X=x)	w spike	LL	Expected	Chisq			
0	12	0.003732	0.03116686	-41.6208009	11.9992393	4.8219E-08	chisq	61.95977612	
1	8	0.00847416	0.0082408	-38.3892612	3.17270864	7.34474689	df	48	
2	8	0.01338169	0.01301319	-34.7343346	5.0100781	1.78433006	pval	0.0849096	
3	7	0.01804652	0.01754956	-28.2990848	6.75658064	0.00876967			
4	8	0.02225746	0.02164454	-30.6640147	8.33314972	0.01331894			
5	10	0.02591175	0.0251982	-36.8098265	9.70130766	0.0091964			
6	10	0.02897138	0.02817357	-35.6937084	10.8468259	0.06611281			
7	5	0.03143822	0.03057248	-17.4382749	11.7704054	3.89437642			
8	8	0.03333862	0.03242056	-27.4317008	12.4819142	1.6093329			
9	10	0.03471348	0.03375755	-33.8855121	12.9966564	0.69094306			
10	8	0.03561155	0.03463089	-26.9040729	13.3328942	2.13305226			
11	12	0.03608509	0.03509139	-40.197593	13.510186	0.16881054			
12	15	0.03618682	0.03519032	-50.2047629	13.5482737	0.15555555			
13	12	0.035968	0.03497753	-40.2365938	13.4663482	0.15967039			
14	16	0.03547717	0.03450021	-53.8686387	13.2825802	0.55594398			
15	12	0.03475937	0.03380217	-40.6467616	13.0138374	0.07898256			
16	15	0.03385577	0.03292346	-51.2035454	12.6755335	0.42626565			
17	7	0.03280351	0.03190018	-24.1160054	12.2815692	2.27128741			
18	12	0.03163568	0.03076451	-41.7767215	11.8443372	0.00204578			
19	13	0.03038149	0.02954486	-45.7839926	11.3747701	0.23221326			
20	16	0.02906644	0.02826602	-57.0575191	10.8824172	2.40660256			
21	14	0.02771259	0.02694945	-50.5930931	10.3755385	1.26612425			
22	9	0.02633883	0.02561353	-32.9817126	9.86120729	0.07521168			
23	9	0.02496118	0.02427381	-33.4652151	9.34541732	0.01276702			
24	6	0.02359305	0.02294335	-22.6483631	8.83318944	0.90872753			
25	17	0.02224551	0.02163292	-65.1701625	8.32867489	9.02807232			
26	6	0.0209276	0.02035131	-23.3676604	7.83525375	0.42987202			
27	6	0.01964654	0.01910553	-23.7466662	7.35562729	0.24983938			
28	7	0.01840796	0.01790105	-28.1602722	6.89190339	0.00169545			
29	3	0.0172161	0.01674201	-12.2695019	6.44567495	1.84196007			
30	9	0.01607405	0.01563141	-37.4262589	6.01809111	1.47750847			
31	11	0.01498384	0.01457122	-46.5157727	5.60992148	5.17885081			
32	2	0.01394669	0.01356263	-8.60087368	5.22161372	1.98766043			
33	5	0.01296306	0.01260609	-21.8678761	4.85334475	0.00443153			
34	6	0.01203283	0.01170147	-26.6882445	4.50506616	0.49606978			
35	8	0.01115536	0.01084817	-36.1900736	4.17654408	3.50021811			
36	6	0.01032963	0.01004518	-27.6039745	3.86739411	1.1759877			
37	6	0.0095543	0.0092912	-28.0721258	3.57711178	1.64109698			
38	3	0.00882777	0.00858467	-14.2733309	3.30509883	0.02816415			
39	3	0.00814824	0.00792386	-14.5136307	3.05068589	0.00084213			
40	3	0.0075138	0.00730689	-14.7568135	2.81315185	0.01241036			
41	3	0.00692242	0.00673179	-15.0027413	2.59174028	0.06431045			
42	3	0.00637202	0.00619655	-15.2512857	2.38567342	0.15819313			
43	0	0.00586051	0.00569913	0	2.19416369	2.19416369			
44	1	0.00538577	0.00523746	-5.251918	2.01642338	0.51235097			
45	1	0.00494573	0.00480954	-5.33715403	1.85167253	0.39172483			
46	3	0.00453834	0.00441336	-16.2693539	1.69914533	0.99592591			
47	2	0.0041616	0.004047	-11.0195586	1.5580952	0.12533243			
48	0	0.00381359	0.00370857	0	1.42779873	1.42779873			
49	0	0.00349243	0.00339626	0	1.30755874	1.30755874			
50+	9	0.03583247	0.03484573	-30.2114227	13.4156052	1.45334994			
	385								

Appendix F: LRT Test Between Spiked and Unspiked Model

LL w/o spike	LL w spike	2 x [LL(big) - LL(small)]			
-57.77634015	-41.6208009	32.31107852	LRT	21.53118412	
-33.62765659	-38.3892612	-9.523209246	df	1	
-31.08277659	-34.7343346	-7.3031161	p-val	3.48121E-06	
-25.81427943	-28.2990848	-4.969610665			
-28.45075127	-30.6640147	-4.426526768			
-34.66853146	-36.8098265	-4.282590031			
-34.06317954	-35.6937084	-3.261057722			
-16.83495487	-17.4382749	-1.206639954			
-26.75096516	-27.4317008	-1.361471261			
-33.33539621	-33.8855121	-1.100231802			
-26.66847304	-26.9040729	-0.471199647			
-40.1055321	-40.197593	-0.184121729			
-50.36815011	-50.2047629	0.326774392			
-40.55690369	-40.2365938	0.640619733			
-54.51007613	-53.8686387	1.282874833			
-41.26331548	-40.6467616	1.233107824			
-52.11562396	-51.2035454	1.824157061			
-24.59600148	-24.1160054	0.959992109			
-42.67490331	-41.7767215	1.796363653			
-46.82119443	-45.7839926	2.07440365			
-58.39360982	-57.0575191	2.67218142			
-51.79877677	-50.5930931	2.411367385			
-33.77122436	-32.9817126	1.579023574			
-34.26080376	-33.4652151	1.591177236			
-23.17767864	-22.6483631	1.058631114			
-66.65341443	-65.1701625	2.966503954			
-23.88095546	-23.3676604	1.026590057			
-24.24567511	-23.7466662	0.998017858			
-28.72134672	-28.1602722	1.122148976			
-12.49914304	-12.2695019	0.459282213			
-38.07773825	-37.4262589	1.302958609			
-47.26054388	-46.5157727	1.489542322			
-8.725949463	-8.60087368	0.25015157			
-22.15242287	-21.8678761	0.569093463			
-26.99332591	-26.6882445	0.610162747			
-36.54506925	-36.1900736	0.709991217			
-27.82905847	-27.6039745	0.450168			
-28.25384059	-28.0721258	0.363429684			
-14.34145664	-14.2733309	0.136251538			
-14.55802967	-14.5136307	0.088798031			
-14.77653884	-14.7568135	0.039450713			
-14.99689085	-15.0027413	-0.011700862			
-15.21899901	-15.2512857	-0.064573345			
0	0	0			
-5.222722103	-5.251918	-0.058391788			
-5.298359951	-5.33715403	-0.077588164			
-16.12345743	-16.2693539	-0.291792859			
-10.90215828	-11.0195586	-0.234800577			
0	0	0			
0	0	0			
-28.21917926	-30.2114227	-3.984486819			