

Characterization of Pathogen- Driven Selection at *B4galnt2* in House Mice

Dissertation

in fulfillment of the requirements for the degree “Dr. rer. nat.”
of the Faculty of Mathematics and Natural Sciences
at Kiel University

Submitted by

Marie Vallier

Kiel, May 2017

First referee: Pr.Dr. John Baines

Second referee: Pr.Dr. Hinrich Schulenburg

Date of the oral examination: 12 May 2017

Approved for publication: 12 May 2017

Zusammenfassung

B4galnt2 ist eine Blutgruppen-assoziierte Glykosyltransferase, die in Hausmäusen *cis*-regulatorische Variation hinsichtlich der gewebsspezifischen Expression aufweist. Das Wildtypallel, das beispielsweise im Laborstamm C57BL/6J auftritt, führt zur Expression von *B4galnt2* im Intestinaltrakt, die auch bei anderen Vertebraten zu finden ist. Eine alternative Klasse, die sich im Stamm RIIS/J und weiteren Mäusen findet, resultiert in der Expression in den Blutgefäßen, die einen Phänotyp zur Folge hat, der dem Typ 1 von-Willebrand-Syndrom ähnelt, die eine häufige Blutgerinnungsstörung bei Menschen darstellt. Vorherige Studien zeigten, dass die unterschiedlichen *B4galnt2*-Allele in der Maus langfristig einer *Balancing Selection* unterliegen und die Variation der *B4galnt2*-Expression die Interaktion zwischen Wirt und Mikroorganismen im Darm beeinflusst. Das impliziert, dass die Nachteile durch eine höhere Blutungsdauer in Mäusen, die das RIIS/J-Allel tragen, durch Vorteile in der Resistenz gegenüber gastrointestinalen Pathogenen aufgewogen werden. Die Bedingungen unter denen ein solcher Ausgleich zur langfristigen Erhaltung krankheits-assoziiierter Variation führt, sind jedoch bislang ungeklärt.

Um die potentiell pathogenbedingten Selektionsmechanismen zu verstehen und charakterisieren, die in der Natur auf *B4galnt2* wirken, habe ich zunächst ein mathematisches Modell entwickelt, das auf einem evolutionären Spiel mit einem modifizierten Wright-Fisher-Prozess basiert und für diploide Individuen angepasst wurde. Dabei habe ich mich insbesondere auf heterozygote Mäuse konzentriert, die *B4galnt2* in Blutgefäßen und Gastrointestinaltrakt exprimieren. Im Vergleich von simulierten und freilebenden Populationen zeigt sich, dass die Frequenzen der Genotypen, die in der Natur beobachtet werden, durch pathogenbedingte Selektion hervorgerufen werden können, falls (i) die Nachteile der Gerinnungsstörung etwa halb so groß sind wie die durch Infektion sowie (ii) sowohl heterozygote als auch das RIIS/J-Allel tragende Individuen gegen Infektionen geschützt sind. Die Resistenz der heterozygoten Individuen zeigt, dass eine dominante protektive Eigenschaft des RIIS/J-Allels wahrscheinlicher ist als eine protektive Wirkung des Verlustes der intestinalen Expression. Der Hintergrund der dominanten protektiven Funktion des RIIS/J-Allels bleibt jedoch unklar, da das Modell impliziert, dass die damit assoziierte vaskuläre Expression nicht unbedingt mit einer Resistenz gegen Pathogene verbunden ist.

Weiterhin habe ich mittels „Deep Phenotyping“ in mehr als 200 neugefangenen Mäusen aus Südfrankreich, wo mittlere Häufigkeiten des RIIS/J-Allels auftreten, auf die Identifikation potentieller Pathogene abgezielt, die die Selektion von *B4galnt2*-Varianten antreiben. Durch die

Analyse von genetischen Mustern, Entzündungszeichen und der intestinalen mikrobiellen Gemeinschaften, konnte ich mehrere Bakteriengattungen mit Mustern in Verbindung bringen, die mit genotypabhängigen Wirt-Pathogen-Interaktionen konsistent sind. Besonders Arten aus der Gattung *Morganella* sind wahrscheinliche Kandidaten, da diese bekannte opportunistische Pathogene einschließt und zudem Häufigkeiten, Prävalenzen und Aktivitätsmuster mit einer erhöhten Entzündung in Mäusen einhergehen, die eine intestinale Expression von *B4galnt2* aufweisen. Zudem konnte ich die entsprechende *Morganella*-Art identifizieren, die eine neue Art der *Morganella morgani*-Gruppe darstellt und über Virulenz-assoziierte Gene verfügt, die nicht in anderen *Morganella*-Arten vorkommen und für ihr Potential über genotypabhängige Wirt-Pathogen-Interaktionen die Selektion von *B4galnt2* voranzutreiben, verantwortlich sein könnten.

Insgesamt hat meine Arbeit dazu beigetragen, neue Einsichten in mögliche evolutionäre Dynamiken von *B4galnt2* in wildlebenden Hausmaus-Populationen zu entwickeln, die zeigen, dass die Pathogen-abhängige Selektion ein wahrscheinlicher Grund für die Erhaltung beider *B4galnt2*-Allele in der Natur ist. Weiterhin bietet meine Arbeit auch über Glykosyltransferasen der Maus hinaus Perspektiven für den Einsatz der hier entwickelten Methoden, die leicht für andere biologische Modelle generalisierbar sind.

Abstract

B4galnt2 is a blood group-related glycosyltransferase that displays cis-regulatory variation for its tissue-specific expression patterns in house mice. The wild type allele, found e.g. in the C57BL/6J laboratory mouse strain, directs intestinal expression of *B4galnt2*, which is the pattern observed among vertebrates, including humans. An alternative allele class found in the RIIS/J strain and other mice instead drives expression in blood vessels, which leads to a phenotype similar to type 1 von Willebrand disease (VWD), a common human bleeding disorder. Previous studies showed that alternative *B4galnt2* alleles are subject to long-term balancing selection in mice and that variation in *B4galnt2* expression influences host-microbe interactions in the intestine. This suggests that the cost of prolonged bleeding in RIIS/J allele-bearing mice might be outweighed by benefits associated with resistance against gastrointestinal pathogens. However, the conditions under which such trade-offs could lead to the long-term maintenance of disease-associated variation at *B4galnt2* are unclear.

To understand and characterize the potential pathogen-driven selection acting on *B4galnt2* in the wild, I first developed a mathematical model based on an evolutionary game framework with a modified Wright-Fisher process, adjusted to implement diploid individuals. In particular, I focused on heterozygous mice, which express *B4galnt2* in both blood vessels and the gastrointestinal tract. By comparing simulated to natural populations, I found that the genotype frequencies observed in nature can be produced by pathogen-driven selection when (i) the fitness cost of bleeding is roughly half that of infection and (ii) both heterozygotes and RIIS/J homozygotes are protected against infection. The resistance of the heterozygote individuals indicates that a dominant protective function of the RIIS/J allele is more likely than a protective loss of intestinal expression. However, the nature of the dominant protective function of the RIIS/J allele remains unknown, as the model suggests that the associated vascular expression is not necessarily linked to the pathogen resistance.

Furthermore, I aimed to identify potential pathogens driving the selection at *B4galnt2* by sampling and phenotyping over 200 newly collected mice from Southern France, where an intermediate frequency of the RIIS/J allele is present. Through the multilayer analysis of genetic patterns, signs of inflammation, and intestinal microbial communities, I could associate several bacterial genera to patterns consistent with genotype-dependent host-pathogen interaction. One genus in particular, *Morganella*, is a likely candidate as it is a well-known opportunistic pathogen and its abundance, prevalence and activity patterns are associated with increased inflammation

in mice with intestinal expression of *B4galnt2*. Finally, I could identify the relevant species of *Morganella*, which represents a new subspecies of the *Morganella morganii* group, and possesses virulence-related genes absent from the other *Morganella* species, which may account for its potential to drive selection at *B4galnt2* via genotype-dependent host-pathogen interactions.

In conclusion, my work provides new insights into the potential evolutionary dynamics taking place at *B4galnt2* in wild populations of house mice, showing that pathogen-driven selection is a likely cause for the maintenance of both *B4galnt2* alleles in the wild. Moreover, my work could be applied beyond the scope of murine glycosyltransferases, as the methods that I developed can easily be generalized to other biological models.