

*DISS. ETH NO...24808...*

***Investigation into the cause of fertility restoration in  
cytoplasmic male sterile perennial ryegrass  
(Lolium perenne L.)***

***A thesis submitted to attain the degree of  
DOCTOR OF SCIENCES of ETH ZURICH  
(Dr. sc. ETH Zurich)***

***presented by***

***TIMOTHY ANDREW SYKES***

***MSc. Biotechnology (Distinction), University of Essex, UK***

***born on 10.08.1984***

***citizen of  
Australia***

***accepted on the recommendation of***

***Prof. Dr. Bruno Studer, examiner***

***Prof. Dr. Torben Asp, co-examiner***

***Prof. Dr. Ian Max Møller, co-examiner***

***2017***

# Summary

Understanding the genetic architecture underlying cytoplasmic male sterility (CMS) is vital if this trait is to be utilised in even more species for the production of high-yielding hybrid varieties. CMS has already been applied to great effect in several agricultural species with the largest yield gains seen in maize, where hybrids now make up to 65% of the global maize production. As global food demand increases, new hybrid varieties, especially in crops where the potential of hybrid breeding is yet to be fully exploited, will play a key role in ensuring that agricultural production can continue to maintain global food security.

The CMS trait is characterised as an inability of affected individuals to produce viable pollen. This phenotype allows plant breeders to use CMS as a pollination control mechanism during the production of hybrid seed. In order to maintain the CMS trait within breeding programs, molecular markers are needed that allow breeders to track genes that can restore male fertility. Although the cause of CMS is inherited through the cytoplasm, the genes that can restore fertility (*Rf* genes) are nuclear encoded. This makes the CMS trait not just agriculturally important for the production of hybrid seed, but also of fundamental interest in the study of mitochondrial-nuclear genome interactions.

This thesis documents an investigation into the cause of CMS restoration in perennial ryegrass (*Lolium perenne* L). Although the importance of forage grasses such as perennial ryegrass is often over-looked, they account for 80% of milk and 70% of meat production in Europe. A CMS system has been identified in perennial ryegrass although it has yet to be fully characterised, making it a challenge to utilise for hybrid seed production. The goal of this research was to identify the genes or genomic loci responsible for fertility restoration and in doing so, to reveal the genetic architecture of CMS fertility-restoration in perennial ryegrass. This was achieved through the use of four contrasting approaches, reported in chapters 2-5, which in concert were able to reveal the genetic control of this important trait.

Firstly, a study of the relevant literature revealed that in almost all cases, identified *Rf* genes belong to a class of RNA binding proteins called the pentatricopeptide repeat (PPR) proteins. Although being particularly numerous in plant genomes, with hundreds of members identified, a sub-class of CMS restoring PPRs, the restorer of fertility-like PPRs (RFLs), has been identified. In **chapter 2**, a bioinformatics pipeline is outlined that allows the rapid identification of these *RFL* genes from genomic or transcriptomic sequence data, exploiting sequence similarities between *RFLs*. This pipeline not only allows the simultaneous identification of *RFLs* from multiple species but also permits the identification of active sites of *RFL* generation within a genome. The application of which, to 14 plant species, revealed that 50-90% of *RFL* genes can be found within two to three distinct genomic loci per species.

A population of over 1,200 perennial ryegrass plants segregating for CMS fertility restoration was identified and formed the basis for the investigation into fertility restoration presented here. This population showed no intermediate phenotypes and a fertility restoration rate of 27%, suggesting that a two loci co-dominant control may be responsible for fertility restoration. Three sequencing-based approaches were employed to interrogate this population, the first of which, genotyping by sequencing (GBS), is presented in **chapter 3**. The application of GBS to this population identified four quantitative trait loci (QTL) associated to the restoration of the CMS phenotype, spanning a genome region of 87.3 Mb.

Following this, a bulk segregant analysis (BSA) was performed on pools of DNA from sterile or fertile plants. As the phenotype of interest is binary, sterile or fertile, this proved to be a powerful way to further resolve the previously identified QTL. The results of the BSA are given in **chapter 4**, where two loci were identified, covering 74 kb of the reference genome. Most strikingly, these two loci were both identified as containing DNA sequence of mitochondrial origin, with one nuclear encoded mitochondrial gene shown to be mutated in the sterile pool. This suggested that functional copies of deficient CMS-causal mitochondrial genes may be responsible for fertility restoration.

Finally, total RNA was collected across three tissue types from the four genotypes present in a CMS breeding program, and a gene expression analysis performed. The results of which, presented in **chapter 5**, revealed several expressed mitochondrial genes within the two QTL identified in chapter 4. Once fully assessed, three strong candidate genes for fertility restoration were identified, two of which were subunits of the mitochondrial respiratory complex IV and have been previously implicated in CMS causation. Mutations in the coding sequence of one of these genes was associated to the sterile phenotype, with both containing predicted mitochondrial transit peptides. In addition to these results, the gene expression analysis uncovered pleiotropic effects associated to the presence of *Rf* genes and highlighted large expression shifts in restored hybrids which suggest a cause for the observations of practical plant breeders that restored hybrids are ‘more vigorous’ than unrestored hybrids.

The results presented here identify a previously undescribed mechanism for CMS fertility restoration and strong targets for marker development within the population studied. Moreover, they add to the body of knowledge concerning nuclear-mitochondrial genome interactions and will be of special interest to researchers in the field of evolutionary gene transfer from mitochondrial to nuclear genomes. The most important outcome of these results is the possibility for the identification and engineering of this novel CMS system in new crops. This will give plant breeders a new approach for integrating CMS into hybrid breeding schemes, and may lead to the development of hybrids in species where the potential of hybrid breeding is yet to be fully realised.

# Zusammenfassung

Das Verständnis der genetischen Architektur, die der cytoplasmatisch männlichen Sterilität (CMS) zugrunde liegt, ist von entscheidender Bedeutung, wenn dieses Merkmal in weiteren Kulturarten für die Produktion von Hohertrags-Hybridsorten genutzt werden soll. CMS wurde bereits bei einer Vielzahl landwirtschaftlicher Arten mit großem Erfolg eingesetzt, wobei der größte Ertragszuwachs bei Mais zu verzeichnen ist, dessen Anteil an Hybriden, gemessen an der weltweiten Maisproduktion 65% ausmacht. Da die weltweite Nachfrage nach Nahrungsmitteln zunimmt, werden neue Hybridsorten, insbesondere in Kulturarten, in denen das Potenzial der Hybridzüchtung noch nicht vollständig ausgeschöpft ist, eine Schlüsselrolle spielen, um sicherzustellen, dass die landwirtschaftliche Produktion weiterhin die globale Ernährungssicherheit aufrechterhalten kann.

Die CMS-Eigenschaft wird durch eine Unfähigkeit der betroffenen Pflanzen charakterisiert, lebensfähigen Pollen zu bilden. Dieser Phänotyp gibt Pflanzenzüchtern die Möglichkeit, CMS als Bestäubungskontrollmechanismus während der Produktion von Hybridsaatgut zu verwenden. Um das CMS-Merkmal innerhalb von Zuchtprogrammen zu verwenden, werden molekulare Marker benötigt, die es Züchtern erlauben, Pflanzenmaterial zu identifizieren, die die männliche Fruchtbarkeit wiederherstellen können. Obwohl die Ursache von CMS durch das Cytoplasma vererbt wird, sind Gene, die die Fertilität wiederherstellen können (*Rf*-Gene), nuklear codiert. Dadurch ist das CMS-Merkmal nicht nur für die Produktion von Hybridsaatgut von agronomischer Bedeutung, sondern auch von grundlegendem Interesse für die Untersuchung von Mitochondrien-Zellkern-Wechselwirkungen.

In dieser Dissertation wird die Untersuchung zur Ursache der CMS-Restauration bei Deutschem Weidelgras (*Lolium perenne* L.) dokumentiert. Obwohl die Bedeutung von Futtergräsern wie Weidelgras häufig vernachlässigt wird, sind sie in Europa für 80% der Milch und 70% der Fleischproduktion verantwortlich. Ein CMS-System konnte in Deutschem Weidelgras identifiziert werden. Da es jedoch noch nicht vollständig charakterisiert wurde, ist dessen Einsatz in der Produktion von Hybrid-Saatgut nur bedingt praktikabel. Das Ziel dieser Untersuchungen war es, die für die Fertilitätsrestauration verantwortlichen Gene oder Genom-Abschnitte zu identifizieren und damit die genetische Architektur der CMS-Fertilitätsrestauration bei Deutschem Weidelgras aufzuklären. Durch die Verwendung von vier verschiedenen Ansätzen, über die in den Kapiteln 2-5 berichtet wird, sowie deren Zusammenspiel, konnte die genetische Kontrolle dieses wichtigen Merkmals aufgezeigt werden.

Zu Beginn der Dissertation zeigte eine Recherche der relevanten Literatur, dass identifizierte *Rf*-Gene in fast allen Fällen zu einer Klasse von RNA-bindenden Proteinen gehören, die als Pentatricopeptid-Repeat (PPR) Proteine bezeichnet werden. Obwohl sie in Pflanzengenomen, mit Hunderten von identifizierten Mitgliedern, besonders zahlreich sind, konnte eine Unterklasse von PPRs identifiziert werden, die CMS wiederherstellt - die restorer of fertility-like PPRs (RFLs). In **Kapitel 2** wird eine Bioinformatik-Pipeline vorgestellt, die die schnelle Identifizierung dieser *RFL*-Gene in genomischen- oder transkriptomischen Sequenzdaten ermöglicht, wobei Sequenzähnlichkeiten zwischen *RFLs* ausgenutzt werden. Diese Pipeline ermöglicht nicht nur die gleichzeitige Identifizierung von *RFLs* mehrerer Spezies, sondern auch die Identifizierung von aktiven Abschnitten im Genom, die zur Neubildung von *RFLs* innerhalb eines Genoms verantwortlich sind. Die Analyse von 14 Pflanzenarten ergab, dass 50-90% der *RFL*-Gene innerhalb von zwei bis drei verschiedenen genomischen Loci pro Art zu finden sind.

Eine in CMS-Fertilitätsrestauration spaltende Population von über 1.200 Weidelgraspflanzen, wurde identifiziert und bildete die Grundlage für die hier vorgestellte Untersuchung zur Wiederherstellung der Fertilität. Diese Population zeigte keine intermediären Phänotypen und eine Wiederherstellungsrate der Fertilität von 27%, was darauf hindeutet, dass eine co-dominante Kontrolle mit zwei Loci für die Wiederherstellung der Fruchtbarkeit verantwortlich sein könnte. Drei sequenzierungsbasierte Ansätze wurden verwendet, um diese Population zu untersuchen, der erste davon, Genotyping by sequencing (GBS), wird in **Kapitel 3** vorgestellt. Die Anwendung von GBS in dieser Population identifizierte vier quantitative trait loci (QTL) im Zusammenhang mit der Wiederherstellung des CMS-Phänotyps mit einer Grösse von insgesamt 87.3 Mb.

Im Anschluss daran wurde eine Bulk Sargeant Analysis (BSA) an DNA-Pools aus sterilen oder fertilen Pflanzen durchgeführt. Da der untersuchte Phänotyp entweder steril oder fruchtbar ist, erwies sich dies als ein effektiver Ansatz, um die zuvor identifizierte QTL weiter aufzulösen. Die Ergebnisse der BSA sind in **Kapitel 4** angegeben, in denen zwei Loci identifiziert wurden, die 74 Kb des Referenzgenoms abdecken. Interessanterweise konnte festgestellt werden, dass diese zwei Loci DNA-Sequenz mitochondrialer Herkunft enthielten, wovon bei einem nuklear codierten mitochondrialen Gen im sterilen Pool gezeigt werden konnte, dass es Mutationen aufweist. Dies legt die Vermutung nahe, dass funktionelle Kopien von defizienten CMS-kausalen mitochondrialen Genen für die Wiederherstellung der Fruchtbarkeit verantwortlich sein könnten.

Letztendlich wurde die Gesamt-RNA von vier in einem CMS-Züchtungsprogramm verwendeter Genotypen, einschliesslich drei verschiedener Gewebetypen in einer Genexpressionsanalyse untersucht. Die Ergebnisse, die in **Kapitel 5** vorgestellt werden, zeigten mehrere exprimierte mitochondriale Gene innerhalb der beiden in Kapitel 4 identifizierten QTL. Nach vollständiger Auswertung wurden drei mögliche Kandidaten-Gene für die Fertilitätswiederherstellung identifiziert, von denen zwei für Untereinheiten des mitochondrialen Atmungskomplexes IV codieren und die schon zuvor in Zusammenhang mit der Ursache von CMS gebracht werden konnten. Mutationen in der kodierenden Sequenz eines dieser Gene sind mit dem sterilen Phänotyp assoziiert, wobei beide Gene errechnete mitochondriale Transitpeptide enthalten. Zusätzlich zu diesen Ergebnissen konnte die Genexpressionsanalyse pleiotrope Effekte aufzeigen, die das Vorhandensein von *Rf*-Genen mit enormen Expressionsverschiebungen in restaurierten Hybriden assoziiert, was die Beobachtung von praktischen Pflanzenzüchtern untermauert, die restaurierte Hybriden im Gegensatz zu nicht-restaurierten Hybriden als "wüchsiger" beschreiben.

Die hier präsentierten Ergebnisse zeigen einen noch nie zuvor beschriebenen Mechanismus für die Wiederherstellung der CMS-Fertilität und identifizieren ideale Ansatzpunkte für die Markerentwicklung innerhalb der untersuchten Population. Darüber hinaus ergänzen sie das Wissen über nuklear-mitochondriale Genom-Interaktionen und sind von besonderem Interesse für Forscher auf dem Gebiet des evolutionären Gentransfers von mitochondrialen zu nuklearen Genomen. Die bedeutendste Erkenntnis dieser Dissertation ist die Möglichkeit zur Identifizierung und Entwicklung dieses neuartigen CMS-Systems in neuen Kulturen. Dies wird Pflanzenzüchtern eine neue Möglichkeit eröffnen, CMS in Hybrid-Züchtungsschemata zu integrieren und dadurch zur Entwicklung von Hybriden in Arten beitragen, bei denen das Potenzial der Hybridzüchtung noch nicht vollständig verwirklicht ist.