THE PHYLOGENETIC HISTORY AND ECOPHYSIOLOGICAL DIVERSITY OF *KALMIA BUXIFOLIA* (SAND-MYRTLE)

A thesis presented to the faculty of the Graduate School of
Western Carolina University in partial fulfillment of the
requirements for the degree of Master of Science in Biology.

By

Ellen Jean Quinlan

Co-Director: Dr. Katherine Gould Mathews
Associate Professor of Biology
Biology Department

Co-Director: Dr. Beverly Collins
Professor of Biology
Biology Department

Committee Members: Dr. Robert Young, Geology; Program for the Study of Developed Shorelines
Dr. James Costa, Biology; Director of Highlands Biological Station

June 2018

ACKNOWLEDGEMENTS

I would like to thank my committee members and directors for their assistance and encouragement throughout this process. In particular, Dr. Mathews for her patience and guidance as I developed new skills, Dr. Collins for her constant wisdom, Dr. Young for his support, and Dr. Costa for his thought-provoking questions and careful editing. They have each pushed me to be a better scientist daily.

I also extend my sincere thanks to the following people and organizations, without whom this thesis would not have been possible: Cait Smith, Tesa Madsen-McQueen, Dr. Emily Gillespie, Dr. Zack Murrell, WCU's Program for the Study of Developed Shorelines, Highlands Biological Station, and the North Carolina Native Plant Society. Lastly, I owe deep gratitude to my family for their unwavering and continued support in all my academic pursuits.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

THE PHYLOGENETIC HISTORY AND ECOPHYSIOLOGICAL DIVERSITY OF *KALMIA BUXIFOLIA*

Ellen Jean Quinlan, M.S.

Western Carolina University (June 2018)

Co-Directors: Dr. Katherine Mathews and Dr. Beverly Collins

Kalmia buxifolia (sand-myrtle, Ericaceae) exhibits a highly disjunct distribution across the high-elevation rock outcrops of the southern Appalachians, upper monadnocks of the North Carolina Piedmont, pine savannas of the Carolina Coastal Plain, and New Jersey Pine Barrens. Although widely separated, these regions share commonalities in their microclimates, particularly nutrient-poor, acidic soils with low soil water retention. Still, *K. buxifolia* coastal populations may use water differently from their mountain counterparts that receive much of their water through daily cloud immersion (fog). Previous studies have found significant, among-region variation in morphology, but no clear pattern of genetic divergence using allozymes in *K. buxifolia*. We hypothesized that there was geographically-based variation in cpDNA haplotypes and drought tolerance between mountain and coastal populations. Additionally, the common hypothesis for rock outcrop floristic divergence predicts that these relationships are relatively young, relics of Pleistocene refugia (<18,000 ybp). The goals of this study were to: 1) Reconstruct the phylogeographic history of *K. buxifolia* and date the divergence from its alpine sister species K. procumbens; and 2) assess the variation in water-use efficiency (WUE) between the mountain and coastal populations of *K. buxifolia* over a season. Dating analysis refutes the rock-outcrop hypothesis as it applies to this species, placing the divergence of *K. buxifolia* and *K. procumbens* mid-Miocene (~14.9 Ma). Haplotype analysis then indicated four potential refugial sites with the most ancient found on Mount LeConte (GSMNP) and three more recent refugia around the Highlands Plateau, Sufolk escarpment, and New Jersey Pine Barrens, as well as points to an Appalachian corridor as the likely Pine Barrens colonization route. Both species and population level divergences within *K. buxifolia* seem to coincide with the major climatic shifts from the mid-Miocene to early-Pliocene. WUE results indicate that plant water-use varies geographically within *K. buxifolia* and that the variation is likely driven by stomatal function rather than morphology. Understanding these phylogenetic and ecophysiological relationships within this species and other similar species is essential when confronted with a once-again changing climate and may be important for developing future management plans.

CHAPTER ONE: INTRODUCTION

Climatic oscillations throughout the last ice age played a significant role in shaping many of the plant and animal distributions we see today. The basic expansion-contraction (EC) model, developed to explain the post-Pleistocene distributions of flora and fauna across Europe, provides a simple template for understanding the impact of climate cycling on species distributions (Taberlet et.al 1998, Hewitt 1999, Provan and Bennett 2008). During periods of glaciation, ranges contract as organisms track suitable habitat towards refugia in more southern or ice-free latitudes. During this process, species unable to shift their ranges fast enough (i.e. insufficient dispersal mechanisms) can go extinct, and populations that do establish at southern latitudes are susceptible to genetic bottlenecks and founder effects (Taberlet et.al 1998, Hewitt 1999, Provan and Bennett 2008). As the climate warms again and ice sheets retreat, species expand back northward, and populations grow in size and extent. This cycle has been repeated many times throughout the Quaternary and can become complex when factors such as niche specialization influence species' establishment patterns (Hewitt 1996, Taberlet et.al 1998, Hewitt 2000, Ikeda and Steguchi 2007, Provan and Bennett 2008).

The EC model has since been used to explain the isolation of many northern species in high-elevation communities of the southern Appalachians (latitude 35° – 37°) (Ramseur 1960, Wiser 1994, Godt et al. 1996, Wiser et.al 1996). During expansion periods, populations are thought to have moved altitudinally up the high mountain slopes, as well as latitudinally. Then, as the climate warmed at the end of the Pleistocene (~ 12,000 years BP), populations remaining at lower elevations in southern latitudes disappeared because they could not acclimate or were pushed out by expanding hardwood forests. This has resulted in a distinctive flora for high-elevation communities, comprised of many rare and disjunct species (e.g. *Sibbaldiopsis tridentata*, *Huperzia appalachiana, Minuartia groenlandica*) (Wiser 1994).

Kalmia buxifolia (Bergius) Gift & Kron (sand-myrtle, Ericaceae) is a long-lived woody, flowering shrub with a highly disjunct distribution. It occurs across high-elevation rock outcrops of western North Carolina, the upper monadnocks of the North Carolina Piedmont, Pine Savannas of the Carolina Coastal Plain, and the New Jersey Pine Barrens. Due to its distribution and high morphological variability, classification of this species has long been up for debate. Sand-myrtle was first described as the single species *Ledum buxifolium* by PJ Bergius (1777), a name also used by Persoon (1805) although he placed it in a subgenus of *Ledum* (*L.* subg. *Leiophyllum* Pers.). Elliott (1817) eventually recognized this subgenus as its own unique genus, renaming the species *Leiophyllum buxifolium* (Bergius) Elliott. Sand-myrtle remained in *Leiophyllum* throughout most of the 20th century, although taxonomists debated splitting the species into 2-3 distinct taxa (as either species or varieties), based largely on its disjunct distribution and the high degree of morphological variability (Table 1; Small 1903, Camp 1938, Wood 1961, Wilbur and Racine 1971). The morphological variability used to distinguish these species relied heavily on leaf size and shape across all treatments, however no author recovered any reliable characters for identification (see key in Strand and Wyatt 1991). Thus, Wilbur and Racine (1961) concluded *Leiophyllum* “should be treated as a not particularly polymorphic monotypic genus and species”. Strand and Wyatt (1991) attempted to answer questions about the evolutionary history of *Leiophyllum* by assessing variation in morphology and allozymes. Although they did find significant among-region variation for seven morphological characters (including alternate vs. opposite leaf arrangement and density of glandular trichomes), allozyme variation was lower than that found in other long-lived woody species and no clear pattern of genetic divergence could be determined. Finally, Kron and King (1996) reclassified the plant as a single species in the genus *Kalmia* based on cladistic analyses of two nucleotide sequences, the chloroplast *rbcL* gene and the nuclear ribosomal internal transcribed spacers (*ITS*) +5.8s region. These sequences indicated that some *Kalmia* species were more closely related to both *Leiophyllum* and *Loiseleuria* than other kalmias, and their inclusion in *Kalmia* made the previously paraphyletic genus monophyletic (Kron and King 1996).

Kalmia procumbens (formerly *Loisleuria procumbens* (L.) Desvaux), now identified as the sister species (Gillespie and Kron 2010), is a circumarctic and alpine-arctic plant- found in the high peaks of New England and northern boreal regions around the world (Ikeda et.al 2017). *K. buxifolia* and *K. procumbens* are united within *Kalmia* as they have the only two flowers in the genus whose petals are not fused the length of the corolla and both lack the anther pockets and tension-loaded pollen dispersal mechanism that characterizes the rest of the genus (Gillespie and Kron 2013). As this sister clade is relatively recently derived in *Kalmia*, the absence of these characters is hypothesized to be a secondary loss (Gillespie and Kron 2013). As *K. procumbens* is the only member of *Kalmia* not restricted to North America and Cuba, the most recent common ancestor (MRCA) of *K. buxifolia* and *K. procumbens* is thought to have occurred at southern latitudes in North America during the late Pliocene/early Pleistocene, with divergence in the expansion period that followed (Ikeda et al. 2017).

Table 1. 20th century taxonomic treatments of *Kalmia buxifolia* by distribution (adapted from Strand and Wyatt 1991).

	Distribution		
Author	**New Jersey Pine Barrens**	**Atlantic Coastal Plain; New Jersey Pine Barrens; Southern Appalachians; Piedmont**	**Southern Appalachians**
Small (1903)	*Dendrium buxifolium* (Bergius) Desvaux	*D. hugeri* Small	*D. prostratum* (Loudon) Small
Camp (1938)	*Leiophyllum buxifolium var. typicum* C. Schneider	*L. buxifolium var. hugeri* (Small) C. Schneider	*L. lyonia (Sweet) Sweet*
Wood (1961)	*Leiophyllum buxifolium var. buxifolium*	*L. buxifolium var. hugeri* (Small) C. Schneider	*L. buxifolium var. prostratum* (Loudon) Gray
Wilbur and Racine (1971)	*Leiophyllum buxifolium* (Bergius) Elliott	*L. buxifolium* (Bergius) Elliott	*L. buxifolium* (Bergius) Elliott
Kron and King (1996)	*Kalmia buxifolia* (Bergius)	*K. buxifolia* (Bergius)	*K. buxifolia* (Bergius)

Camp (1938) hypothesized *K. buxifolia* originated in what is now the North Carolina Piedmont during the Late Mesozoic, based upon its current distribution and that of other members of the tribe Phyllodoceae Drude (Ericoideae, Ericaceae) endemic to the southeastern United States. Historically, both *Leiophyllum* and *Loiseleuria* were included in Phyllodoceae, before being reclassified into *Kalmia*, one of the seven remaining genera in the tribe (*Bejaria, Elliottia, Epigaea, Kalmia, Kalmiopsis*, *Phyllodoce*, and *Rhodothamnus*). Camp suggested the species began migrating during the Early Tertiary in two directions – east and north up the Coastal Plain where submergences in the Late Tertiary separated populations in the New Jersey Pine Barrens from the Carolina coast, and west towards the Appalachian Mountains. Finally, climatic oscillations during the Pliocene and Pleistocene reduced the available suitable mountain habitat and fragmented the population system. Strand and Wyatt (1991) pointed to issues with this theory as it pertains to the northern migration and coastal separation, as ice sheets of the Pleistocene reached south of New Jersey after the predicted colonization. They asserted that such ice would have killed off any populations connecting the coastal regions at that time. However, although ice extended as far south as New Jersey, the sandy soils and marine climate tempered by the nearby Atlantic Ocean maintained an ice-free periglacial zone in the Pine Barrens (French et.al 2003). This is supported by pollen sampling of the Pine Barrens, which shows a very different vegetation history in comparison to glaciated parts of New Jersey and aids the hypothesis that this region served as a refugium for both northern and southern affinity species during the Pleistocene (Potzger 1945). The Pine Barrens are also structurally high (at least 10-30m above sea-level) and would not have been submerged during most Pleistocene flood cycling (Stanford 2015). During the height of the last glaciation (~21,000 ya), global sea-level dropped to about 125m lower than it is today (USGS 2012). The exposed continental shelf along eastern North America would have become a sandy, pine-forest corridor, able to connect the outer edges of the Carolinas to New Jersey. Climatic oscillations throughout the Pleistocene resulted in rise and fall of sea-level at least every 100,000 years, and thus the corridor may have been open several times for a few thousand years at time, allowing for much exchange along the east coast (Pisias and Moore 1981).

However, the exposed eastern coastline is not the only corridor that may have allowed *K. buxifolia* to colonize New Jersey. Historically, *K. buxifolia* was documented in eastern Kentucky (Whitely Co.; observations and herbarium specimens) and eastern Pennsylvania (Monroe Co., and Northampton Co.; observations only, now extirpated), suggesting a once more robust and interconnected Appalachian population that has since been extirpated (Pennsylvania Natural Diversity Inventory, Medley 1993, Rhoads and Klein 1993, Thomas 1997, Campbell and Medley 2012, Kartesz 2014). Thus, several methods for colonization of the Pine Barrens could be hypothesized. First, Pine Barrens populations may be relicts of the ancient arcto-tertiary flora that were never connected with southern populations and persisted through the ice age in their current location. This hypothesis would also suggest that the species is much older than currently thought (at least pre-Pleistocene). Alternatively, either the coastal corridor opened during a drop in sea-level or the Appalachians served as a corridor of cooler, rocky habitat that provided for species exchange and colonization between the Pine Barrens and the Carolinas. Supporting this hypothesis is the proximity of the historical Pennsylvania and New Jersey populations, with just around 100 miles between them. Jump dispersal is unlikely as fruits are small, dehisce before abscission, and lack eliasomes or any other unusual features, making gravity the most likely dispersal mechanism (Strand and Wyatt 1991). However, had there once been suitable habitat connecting northeastern Pennsylvania and the Pine Barrens, it may have supported species exchange between them.

The current distribution of *K. buxifolia* populations throughout the Piedmont and Coastal Plain of the Carolinas also provides clues to past migration routes. Almost linearly, populations through the Sandhills demarcate the Sufolk escarpment (a Pleistocene sea-level highstand shoreline) which, like the Pine Barrens, is structurally high (14-75m above sea level) and would have been the coastline during the highest flooding events (Divins and Metzger 2008, Parham et.al 2012). This pattern suggests *K. buxifolia* could have migrated along that ancient coastline or that populations east of the escarpment were wiped out during high ocean periods and recently recolonized.

Historically, migratory patterns of plants following climatic oscillation have been inferred from the fossil record; more recently, phylogeographic studies have used molecular markers to identify refugia from glacial periods and track recolonization (Stehlik et al. 2001, Petit et al. 2003, Dorken and Barrett 2004, Soltis et al. 2006, Ikeda et al. 2007, Smith et al. 2011). Applying the EC model, refugia are expected in the southernmost part of a distribution and retain the greatest diversity, while recolonized populations retain part of the genetic structure of their refugial ancestors (Ikeda and Setoguchi 2006). Chloroplast DNA (cpDNA) can be particularly useful in locating refugia and tracing post-glacial colonization in plants, as cpDNA is typically maternally inherited, which greatly reduces the effective population size relative to nuclear, and thus retains finer geographical structure (Vendramin et.al 1999, Petit et.al 2002, Ikeda and Setoguchi 2006). Additionally, cpDNA's non-recombining characteristics can preserve haplotypes over generations with relatively few changes among sites (King and Ferris 1998, Petit et.al 2002, Ikeda and Setoguchi 2007). Non-coding and spacer regions in both nuclear and cpDNA are particularly helpful in tracing migrations and assessing relationships within a genus or among populations due to their high mutation rates (White et.al 1990). Gillespie and Kron (2013) found a combination of nuclear and cpDNA gene regions, both coding and non-coding (rbcL, *mat*K, ndhF, *trn*S-G-G, nrITS, and WAXY), sufficient for resolving the relationships within the Ericaceae tribe Phyllodoceae of which the genus *Kalmia* is a member. Phylogeographic studies of the sister species (*K. procumbens*) and other Phyllodoce members have also found the cpDNA *trn*L-F spacer to be highly variable (Taberlet et.al 1991, Ikeda and Setoguchi 2006, Ikeda et.al 2017). Thus, regions found to be variable in these studies should be expected to be sufficiently variable within *K. buxifolia* to infer phylogeography.

Understanding the phylogeographic relationships within *K. buxifolia* and its divergence from *K. procumbens* may be key in predicting how the species will respond to current and future climatic pressures. In the southern Appalachians, *K. buxifolia* inhabits high-elevation rock outcrops, systems unique in their environmental characteristics and community make-up. Outcrops are fairly uncommon on the landscape and are isolated from each other, which contributes to the high occurrence of rare and endemic species in their communities (Wiser et al. 1996). More than 17% of outcrop flora are northern-affinity endemics or disjunct taxa, hypothesized to be remnants of glacial tundra or alpine vegetation whose ranges have retracted in post-Pleistocene warming (Ramseur 1960, Wiser 1994, Godt et al. 1996, Wiser et al. 1996, Wiser 1998). Typical rock outcrops have shallow, acidic soils, and limited water retention and availability – resulting in water stress and limited plant growth (Culatta and Horton 2014). Frequent cloud immersion (fog) may be essential to ameliorating this stress by decreasing the leaf-to-air vapor pressure deficit (VPD), which reduces water loss through transpiration and increases plant water-use efficiency (WUE) (Culatta and Horton 2014). Direct foliar uptake of fog has been documented in other species that experience daily cloud immersion and may be an important source of moisture for *K. buxifolia* (Limm and Dawson 2010, Culatta and Horton 2014, Berry and Smith 2014). Rock outcrops are also high-light environments and cloud immersion may be important for reducing insolation, which can limit plant productivity when soil moisture is low (Horton and Culatta 2016).

Rock outcrops on Mt. Mitchell (2,030m) were documented by Berry and Smith (2012) to be cloud immersed a minimum of 2hrs for 61% of days during the growing season, and results throughout the Blue Ridge region showed a similar trend (1-2hrs immersion on the majority of days). However, the Blue Ridge Province is consistently predicted to experience a rise in average annual temperatures of at least 2.3°C to 3.7°C by 2100 and although precipitation forecasts are more variable, they generally predict either historic or lower levels (2mm to 72mm decrease) and greater seasonal variability over the same period (McNulty et al. 2012, Wear and Greis 2012). Thus, if precipitation does decrease, cloud base height is predicted increase, cloud immersion is predicted to decrease, and growing seasons are predicted to become longer (Richardson et al. 2003). In the southern Appalachians, cloud base height increased over three decades (1970-2003), following the overall global trend, especially after the 1970s (Chernykh et al. 2001, Foster 2001, Richardson et al. 2003, Sun et.al 2007). If periods of rain, fog, or cloud moisture decrease as they often do in summer months, outcrops can quickly become hot and dry and experience drought-like conditions. If growing seasons lengthen, increasing the possibility of drought, and cloud immersion significantly declines or disappears from these slopes altogether, outcrops may become un-inhabitable for many species.

Although widely separated, the coastal sites on which *K.buxifolia* is found share many commonalities with their mountain counterparts. These commonalities include poor soil water retention and nutrient poor, acidic soils. As Coastal Plain soils drain water quickly, sites frequently alternate between saturation and prolonged drought which intensifies in summer months and may become more frequent with climate change. Although all populations of *K. buxifolia* are somewhat drought-tolerant as indicated by the soils on which they live, it is reasonable to predict that Coastal populations may use water differently from their Mountain counterparts, which are adapted to daily cloud immersion and possible foliar uptake. Water-use efficiency (WUE) or the ratio of net photosynthesis to transpiration (A/E) can be useful in assessing water use and drought stress in individuals. As water becomes limited and stress rises, individuals often close their stomates to reduce water loss, which additionally reduces CO_2 uptake and net photosynthesis. Short-term, or leaf-level WUE can be measured directly as the ratio of instantaneous net photosynthesis (A) to transpiration (E). Seasonal WUE frequently is inferred from carbon-isotope discrimination in the leaves (CID), or the ratio of $\delta^{13}C$ to $\delta^{12}C$, as CID is inversely correlated with WUE (Michener and Lajtha 2007). CID is particularly useful in estimating WUE due to its long integration time (time over which carbon in the plant was fixed), which can be up to several months or over seasons (Michener and Lajtha 2007). Variation in WUE may also be apparent in functional traits such as leaf stomata, where plants with ample water will often have high stomatal density that allows them to capitalize on photosynthesis. In comparison, water-stressed plants will have a lower stomatal density to reduce water loss (Milburn and Weatherley 1971). However, agreement between the observed spatial patterns of morphological characters (i.e. stomatal density) and physiological characters (i.e. WUE) is not necessarily expected as they may be under different selection pressures (Ferguson 1980, Strand and Wyatt 1991).

I hypothesize that there is geographically-based variation in cpDNA haplotypes and drought tolerance between mountain and coastal populations of *K. buxifolia* as a result of their phylogeographic history and differing environmental selection on traits related to water use. The goals of this study are to: 1) reconstruct the phylogeographic history of *K. buxifolia* to determine the locations of glacial refugia, directions of postglacial colonization, and divergence time from the sister species, *K. procumbens*; and 2) assess the variation in water-use efficiency (WUE) between the Mountain and Coastal populations of *K. buxifolia* over a season. Here, I test the EC hypothesis that range contraction and expansion during and after the Pleistocene has left scattered remnant populations on high-elevation rock outcrops, along the Sufolk escarpment, and in Coastal sites. I also test whether differences in traits related to WUE indicated the potential for these plants to adapt to climate warming effects on growing season length and longer summer drought. Such adaptation potential has already been documented in other characteristic rock outcrop species such as *Hydatica petiolaris* (cliff saxifrage) and *Solidago simulans* (granitic dome goldenrod), which are able to acclimate to decreasing cloud immersion through decreased transpiration and increased WUE (Horton and Culatta 2016). Understanding this potential in *K. buxifolia* is particularly important if the mountain populations are in fact a glacial refugium, and thus may retain the genetic diversity to select for coast-like drought tolerances as the climate changes. Further, understanding the divergence patterns of *K. buxifolia* may help to characterize other rock-outcrop species with similar mountain-coastal disjunct distributions, such as *Zigadenus leimanthoides* A. Gray (pine barren deathcamas).

CHAPTER TWO: METHODS

Sample Collection

Table 2. *Kalmia buxifolia* sampling localities, analyses applied to each, and voucher ID (samples per population = n).

Collection Site	4-Digit ID	County	Latitude	Longitude	Region	CID	*trn*L-F	*trn*S-G	*nr*ITS	rps4
Cheraw State Park	CHSP	Chesterfield Co., SC	34°37.287'	079°56.678'	Sandhills/Coastal Plain	n=5	n=3	n=2	n=1	n=1
Gordon Butler Nature Preserve	GBNP	Cumberland Co., SC	34°58.989'	078°57.221'	Sandhills/Coastal Plain	n=5	n=3	n=3	n=1	n=1
Grandfather Mountain Inc.	GFMT	Caldwell/Watauga Co., NC	36°05.737'	081°49.918'	Mountains	n=5	n=3	n=2	n=1	n=1
Hanging Rock State Park	HSRP	Stokes Co., NC	36°24.937'	080°16.123'	Piedmont	n=5	n=3	n=3	-	n=1
Laurel Knob	LAKN	Jackson Co., NC	35°09.23880'	083°03.31080'	Mountains	n=5	n=2	n=1	n=1	n=1
Little Green Mountain	LGMT	Jackson Co., NC	35°09.814'	083°00.818'	Mountains	n=5	n=3	n=3	-	n=1
Mount LeConte	MTLC	Sevier Co., NC	35°39.247'	083°26.646'	Mountains	n=5	n=3	n=3	n=1	n=1
Orton Creek	ORCR	Brunswick Co., NC	33°58.851'	078°03.604'	Sandhills/Coastal Plain	n=5	n=3	n=2	-	n=1
Oswego River Preserve	OSRP	Burlington Co., NJ	39°45.824'	074°26.549'	Pine Barrens	n=5	n=3	n=3	-	n=1
Peachtree Rock Heritage Preserve	PRHP	Lexington Co., SC	33°49.993'	081°11.800'	Sandhills/Coastal Plain	n=5	n=3	n=2	-	n=1
Roan Mountain	ROMT	Mitchell Co., NC	36°05.836'	082°08.279'	Mountains	n=5	n=3	n=2	n=1	n=1
Satulah Mountain	SLMT	Macon Co., NC	35°02.185'	083°11.540'	Mountains	n=5	n=3	n=3	n=1	n=1
Sunset Rock	SURK	Macon Co., NC	35°02.790'	083°11.238'	Mountains	n=5	n=3	n=2	-	n=1
Wharton State Forest	WHSF	Burlington/Camden/Atlantic Co., NJ	39°47.619'	074°36.221'	Pine Barrens	n=5	n=3	n=3	-	n=1
Whiteside Mountain	WHMT	Jackson Co., NC	35°04.734'	083°08.468'	Mountains	n=5	n=3	n=3	n=1	n=1

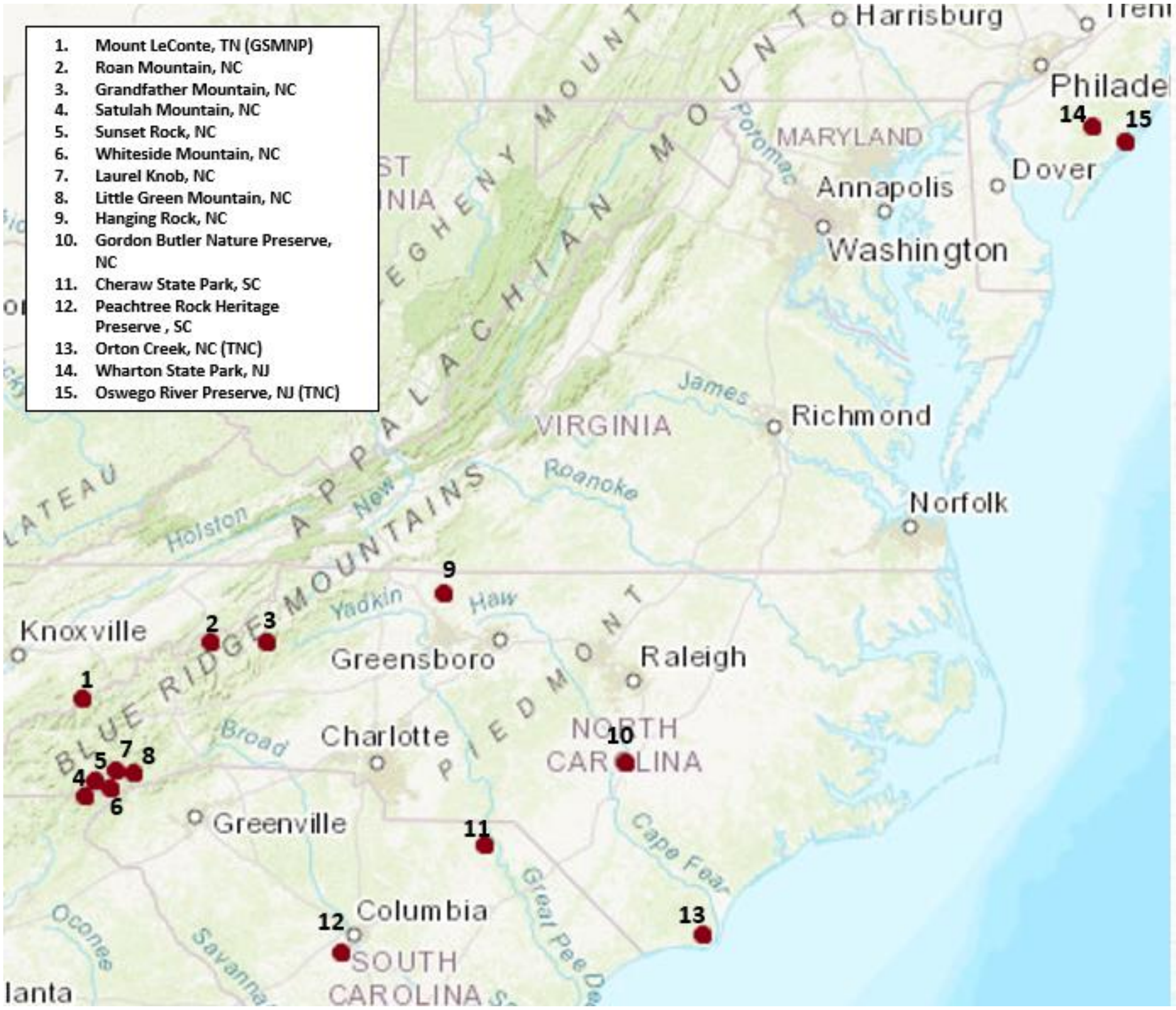


Figure 1. Map of *Kalmia buxifolia* sampling localities.

Study sites (Figure 1; Table 2) were selected to best represent the distribution of *K. buxifolia*, and generally followed the sampling strategy of Strand and Wyatt (1991). Two sets of samples were collected at each site at the start and end of the growing season, the first in May/June and the second in late September/October. Both sets of samples were used in WUE analysis, but DNA was extracted only from those collected in May/June. At each site and in each sampling set, ~5cm clippings were taken from 5-10 individuals, placed into 15mL centrifuge tubes, and dried using Silica gel desiccant, -3+8 Mesh Granules (Alfa Aesar, Ward Hill, MA). Ten leaves were also collected from each site for stomatal analysis.

Leaves chosen were near the ends of branches, fully exposed, and mature. The underside of each leaf was painted with clear, quick-drying nail polish, and allowed to dry before being placed in a resealable plastic bag. Soil depth and moss mat depth (if present) were also measured in three spots, triangulated around the canopy of the individual plant and then averaged. Individuals were also given a “health” rating from 1-5 based on the percent of the canopy that was browned or appeared dead. Individuals with no visible browning were rated at a 5 while individuals with ≥90% of the canopy browning or dead were rated a 1.

DNA Isolation, Amplification, and Sequencing

Silica-dried leaf tissue from three individuals per population was ground to powder in liquid nitrogen using a mortar and pestle. DNA was then extracted and isolated via the DNeasy Plant Mini Kit (Qiagen, Valencia, California) following the manufacturer’s protocol. Four DNA regions were sequenced, both nuclear and chloroplast. The nuclear ribosomal internal-transcribed spacer (*nr*ITS) was PCR-amplified in its entirety using the external primer pair “ITS4-5” (White et.al 1990). Three chloroplast non-coding regions were PCR-amplified, including the *trnL-F* intergenic spacer using primers c and d of Taberlet et al. (1991), the *trnS-G* spacer using primers “*trn*S (GCU)” and “*trn*G(UUC)” of Shaw et al. (2007), and the *rpS4-trnT* intergenic spacer using primers “*rp*S4R2” and “*trn*T(UGU)R” of Shaw et al. (2005). These regions were chosen based on previous sequencing done in *K. buxifolia* and reported high levels of variation in Shaw et al. (2005, 2007). Preliminary samples were sequenced bi-directionally using the same primers as for PCR, but due to the lack of variation found in all markers, remaining samples were sequenced in the single, forward direction using “ITS 4”, “c”, “*trn*S (GCU)” and “*rp*S4R2”. Twenty-five microliter PCR reactions for *nr*ITS were prepared using 12.5µL PCR Taq Mixture (Omega Bio-tek Inc., Norcross, GA), 0.5 µL forward primer (10 µM), 0.5 µL reverse primer (10 µM), 1µL DNA template, and ddH_2O to fill. PCR reactions for cpDNA regions were prepared with 12.5 µL 2x Platinum^{TM} SuperFi^{TM} PCR Master Mix, 1 µL SuperFi^{TM} GC Enhancer (Invitrogen^{TM}, Thermo Fischer Scientific Corp., Carlsbad, CA), 1.25 µL forward primer (10 µM), 1.25 µL reverse primer (10 µM), 5 µL DNA template, and ddH$_2$O to fill to 25 µL.

The ITS region was amplified as follows: initial denaturation cyle at 95^{o}C for 3 minutes; 30 cycles of denaturation at 95 ^{o}C for 45 seconds, annealing at 52 ^{o}C for 45 seconds, and extension at 72 ^{o}C for 2 minutes; followed by a final extension at 72 ^{o}C for 5minutes. Amplifications for the *trn*L-*trn*F intergenic spacer region were performed as follows: initial denaturation at 95 ^{o}C for 2 minutes; 35 cycles of denaturation at 95 ^{o}C for 50 seconds, annealing at 50 ^{o}C for 50 seconds, and extension at 72 ^{o}C for 1 minute 50 seconds; followed by a final extension at 72 ^{o}C for 7 minutes. Amplifications for the *trn*S-*trn*G intergenic spacer region were performed as follows: initial denaturation at 80 ^{o}C for 5 minutes; 30 cycles of denaturation at 95 ^{o}C for 1 minute, annealing at 65 ^{o}C for 1 minute, and extension at 66 ^{o}C for 3 minutes; followed by a final extension at 72 ^{o}C for 2 minutes. The *rp*S4 gene was also amplified as follows: initial denaturation at 98 ^{o}C for 30 seconds; 30 cycles of denaturation at 98 ^{o}C for 10 seconds, annealing at 50 ^{o}C for 30 seconds, and extension at 72 ^{o}C for 30 seconds; followed by a final extension at 72 ^{o}C for 1 minute.

PCR products were run on a 1% w/v agarose gel with a DNA ladder for sizing, then cleaned and prepared for sequencing using the ExoSAP-IT enzymatic cleaner (Thermo Fisher Scientific, Santa Clara, CA). PCR primers were diluted to 5µM and 15 µL sequencing reactions were prepared using one of the following formulas, depending upon band-strength. Samples with strong bands were prepared with 5µL primer, 2 µL PCR product (*trn*S-G) or 1 µL PCR product (ITS, *trn*L-F, *rp*S4-*trn*T), and ddH$_2$O to fill. Samples with weak bands were prepared with 5µL primer, 4 µL PCR product (trnS-G) or 2 µL PCR product (ITS, *trn*L-F, *rp*S4-*trn*T), and ddH$_2$O to fill. Prepared samples were then sequenced by the external vendor GeneWiz (South Plainfield, NJ).

Sequence Alignments and Phylogenetic Analyses

Resulting primer sequences for each sample were downloaded into and viewed in Sequencher (GeneCodes Corp, Ann Arbor, MI) to compare and confirm sequences, and samples with both forward and reverse primer sequences were combined into a consensus. Outgroup sequences for *Phyllodoce nipponica* Makino, *Phyllodoce aleutica* (Spreng.) A. Heller, *Kalmiopsis leachiana* (Henderson) Rehd, *Kalmia latifolia* L. and *Kalmia procumbens* (L.) Gift & Kron were obtained for *nr*ITS, *trn*S-G, and *trn*L-F (*Phyllodoce* sp. unavailable) from GenBank (NIH, Bethesda, MD) (Table 3). Outgroup sequences were then combined with our sequences, aligned in Sequencher, and a Nexus file generated. FindModel (Posada and Crandall 1998) was used to identify the appropriate nucleotide substitution model for each gene region, as determined by the AIC criterion (Akaike 1974). Models were identified for each region as follows: *nr*ITS – GTR + Γ; *trn*S-G – GTR; *trn*L-F – F81; *rp*S4 – F81. As no variation was found among any of our *K. buxifolia nr*ITS sequences, this region was excluded from further analyses.

Divergence dating for *K. buxifolia* and *K. procumbens* was performed using the Bayesian analysis program BEAST v2.4.8 with sequences for the *trn*L-F and *trn*S-G gene regions (Bouckaert et al. 2014). Site model parameters were set and Nexus file converted to a BEAST XML file in BEAUti with partitions for each gene region (Bouckaert et al. 2015). The clock model was set to "Relaxed Clock Log Normal" and the *Kalmia* crown node calibrated using the fossil *Kalmia saxonica,* dated to 15.97-125myr (Mai 2001). The calibration prior used a log-normal distribution and the offset was set as a hard-minimum constraint at 15.97 (Ho 2007). The M and S variables were then set as 54.5 and 2.0 respectively to achieve a mean of 70.5 and 95% CI. Markov-chain monte-carlo (MCMC) analyses consisted of 10,000,000 generations (Yang and Rannala 1997). The BEAST output was then assessed in Tracer v1.6 (Rambaut et al. 2013) and trees were drawn in FigTree v1.4.3 (http://tree.bio.ed.ac.uk/). Clade support is reported as posterior probabilities.

Phylogeographic relationships were analyzed using both TCS (which applies statistical parsimony for network estimation, Clement et al. 2000) and MrBayes (Bayesian inference software) (Huelsenbeck, J. P. and F. Ronquist. 2001). Sequences from genetic markers *trn*L-F, *trn*S-G, and *rp*S4 were used in both analyses with *K. procumbens* as the outgroup. TCS was run twice, once treating gaps as a fifth character state and once treating gaps as missing data; results were the same for both analyses (see below). For the Bayesian phylogenetic analysis, gaps were treated as missing data, since most gaps were phylogenetically uninformative, or in two cases, located within poly-A strings and ambiguous. Bayesian inference of phylogeny among populations was performed in MrBayes, run remotely on the CIPRES Science Gateway online server (v. 3.1) (Miller et al. 2010). As different site models were identified for these regions (*trn*S-G – GTR; *trn*L-F and *rp*S4 – F81) two runs of MrBayes were performed with a different model applied in each for comparison. MCMC was set to 6,000,000 generations with two simultaneous runs and four chains per run. The percentage of trees discarded as burn-in was 0.25. To assess stationarity of likelihood at the end of each run, the value of potential scale reduction factor (PSRF) was confirmed to be approaching 1.0 with a standard deviation of split frequencies approaching 0.0. The ESS statistic was confirmed to be above 100, indicating that the tree sampling from the likelihood distribution was adequate, and the overlay plots of generation vs. log likelihood values showed stationarity for both runs.

Table 3. GenBank accessions for outgroup taxa.

Species	***trn*S-G**	***trn*L-F**	***rp*S4-*trn*T**
Phyllodoce nipponica	GU176702.1	-	-
Phyllodoce aleutica	HM182041.1	-	-
Kalmiopsis leachiana	GU176699.1	KU350274.1	–
Kalmia latifolia	GU176697.1	AB247965.1	–
Kalmia procumbens	JX890296.1	AB247963.1	–

Stable-Isotope Analysis

Five individuals were selected from each site to represent the range in the Carolinas from mountains to Coast. The pair of samples collected at the start and end of the growing season from each individual were prepped for stable-isotope analysis. Samples were oven-dried at 55°C for at least 48hrs and up to 96hrs, then ground to a homogeneous powder using a Mini-Beadbeater-8 (MBB-8) (Biospec Products Inc., Bartlesville, OK). Leaf-tissue was loaded into 2ml screw-cap microtubes with o-ring seals, along with three glass beads (2.5mm). The MBB-8 was then set to 'Homogenize' and run for 2min.

After samples were ground to a powder they were encapsulated and weighed. Tin capsules (4.75x11mm) were weighed on a micro-balance accurate to 0.001mg, tared, and then loaded with sample to 0.7 – 1.5mg. Capsules were then sealed and crimped using forceps down to a spherical package (<2mm), and dropped to confirm there was no leakage. All samples were loaded into a 96-well microtiter plate and sent to the Stable Isotope Analysis Lab in the College of Marine Sciences at the University of South Florida (St. Petersburg, FL) and analyzed for dC13 ($\delta^{13}C$) in a continuous flow isotope ratio mass spectrometer. The carbon isotope composition (δ) was calculated relative to the international Pee Dee Belemnite (PDB) standard (Farquhar et al. 1989) as $\delta_{plant} = (R_{sa} - R_{sd})/R_{sd} \times 1000$, where R_{sa} is the $^{13}C/^{12}C$ ratio of the sample and R_{sd} is that of the standard. Carbon isotope discrimination (CID) was then estimated as $CID = (\delta_{air} - \delta_{plant})/(1 + \delta_{plant}/1000)$, where $-8.0‰ = \delta_{air}$ (^{13}C composition of atmospheric CO_2) (Farquhar et al. 1989, Adriegjo et al. 2014).

Principal Components Analysis (PCA) was performed to examine the association of CID results with environmental variables (soil depth, moss mat, and "health" rating). Samples were identified by geographical region as either "Mountains" or "Piedmont/Coastal Plain" (Table 2). Analysis was performed in PC-ORD™ 7 (McCune and Mefford 2016).

Stomatal Analysis

Stomatal density was assessed as an ancillary method for determining variation in transpiration across regions and compared to results from CID. Leaf peels were generated from the 10 samples collected from each of the 15 sites. In the lab, the nail polish layer on the underside of the leaf was over-laid with clear double-sided tape, then carefully peeled off and mounted on a slide with an imprint of the lower epidermis. Using a compound microscope at 400x, the total number of stomates in the view field were counted in three different areas and then averaged for the approximate stomatal density of the leaf. The radial length and width of each leaf imprint was then measured from the center to compare density with leaf area. Area of the leaf was then calculated as an ellipse using the equation $A = \pi ab$, where area (A) is the product of π, the longest radial length (a), and the longest radial width (b). ANOVAs were then used to test for significant mean differences between regions in stomatal density and leaf area in SAS (SAS Institute Inc. 2011). Significant differences between means ($p<0.05$) were determined using a post-hoc Bonferroni t-test. Finally, linear regression analysis was performed to test for a relationship between stomatal density and CID.

CHAPTER THREE: RESULTS

Divergence Dating Analysis

Divergence analysis in BEAST with the crown node of *Kalmia* calibrated with *Kalmia saxonica* (Mai 2001) dated the divergence of *K. buxifolia* and *K. procumbens* to 14.93Ma within a 95% HPD of 3.5 – 28Ma (Figure 2). The wide HPD interval was the result of the uncertainty around the calibrating fossil date (15.97 – 125 Ma), which was treated conservatively during analysis by placing a hard-minimum constraint on the crown calibration (15.97 Ma) but a relatively soft maximum. This analysis places the divergence of these two species mid-Miocene (~5 – 24 Ma), with the tails of the estimate falling very early Pliocene to late Oligocene.

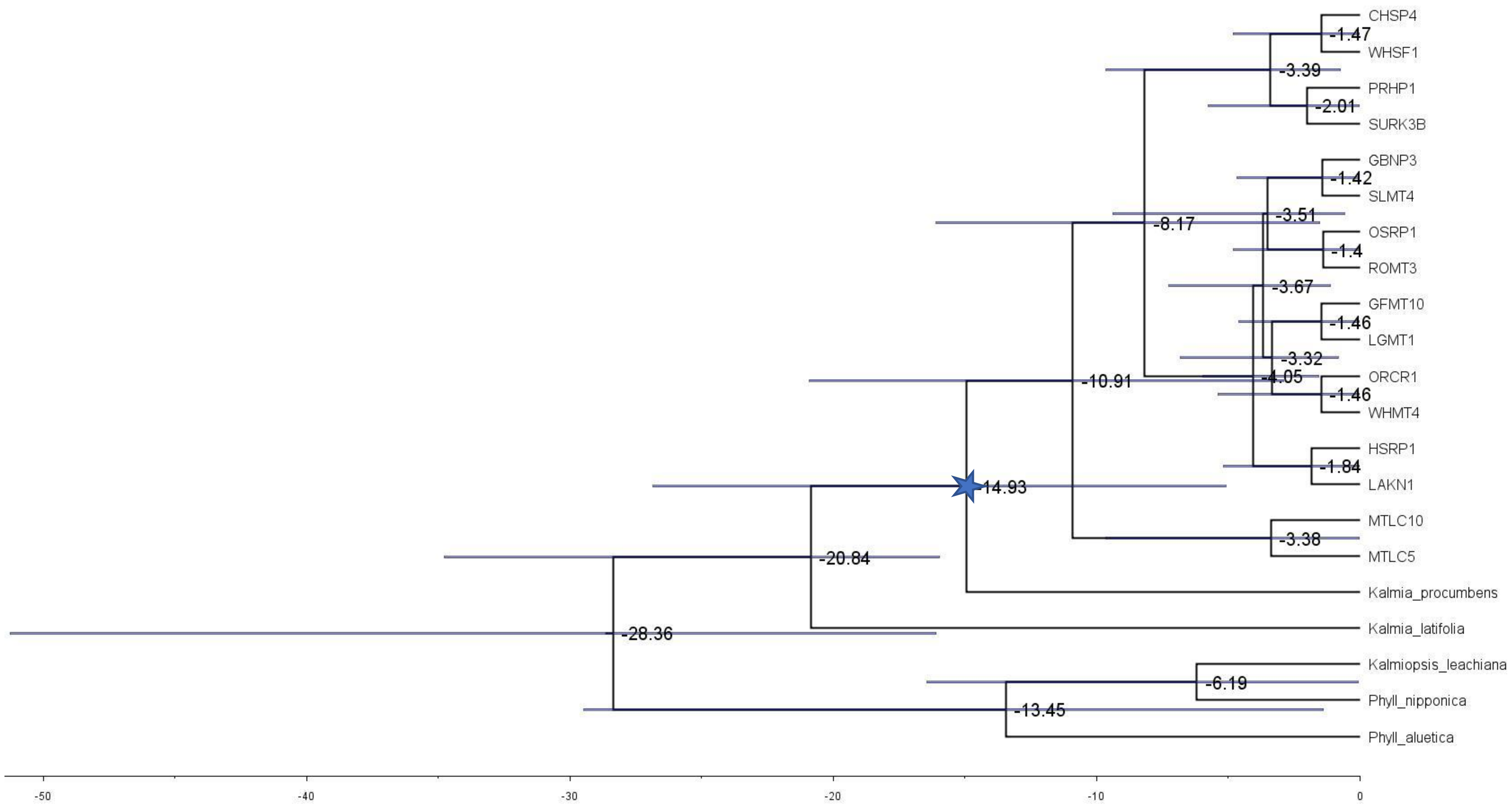


Figure 2. BEAST analysis tree, dating divergence of *Kalmia buxifolia* and *K. procumbens*. Node labels show ages, purple bars over nodes represent the 95% HPD. Divergence of these sister species dated to to 14.93Ma (95% HPD= ~3.5-28Ma).

Phylogenetic Analysis

Analysis of haplotype diversity and distribution in TCS resulted in a network of 10 haplotypes, with haplotype G containing the majority of samples (Figure 3). Haplotype J (followed by H and I) is the most ancestral type relative to the outgroup *K. procumbens* (not shown in Figure 3), which is connected to the H haplotype by 20 mutations. The Mount LeConte population (in GSMNP) contained the greatest haplotype diversity (3 haplotypes) as well as the most ancestral type (J), suggesting it is a refugial population. The three other mountain populations that contain more than one haplotype (Satulah, Laurel Knob, and Little Green Mountain) are all related to at least one of the types found on Mount LeConte, with two of the three (H - found on Laurel Knob, I - found on Satulah) likely derivatives of the most ancestral type (J). All but one of the coastal populations (Peachtree, Gordon Butler, and Orton Creek) also had two haplotypes, with Gordon Butler identical to those found on Laurel Knob and Orton Creek's type E one step away from type f found at Peachtree. Both New Jersey populations (Oswego River and Wharton Creek) exhibited the common haplotype (G), which was found within all regions and all but one population (Mount LeConte). The Oswego River population also contained a second haplotype (A), which is the most derived haplotype in the network and likely a derivative of the mountain haplotype D.

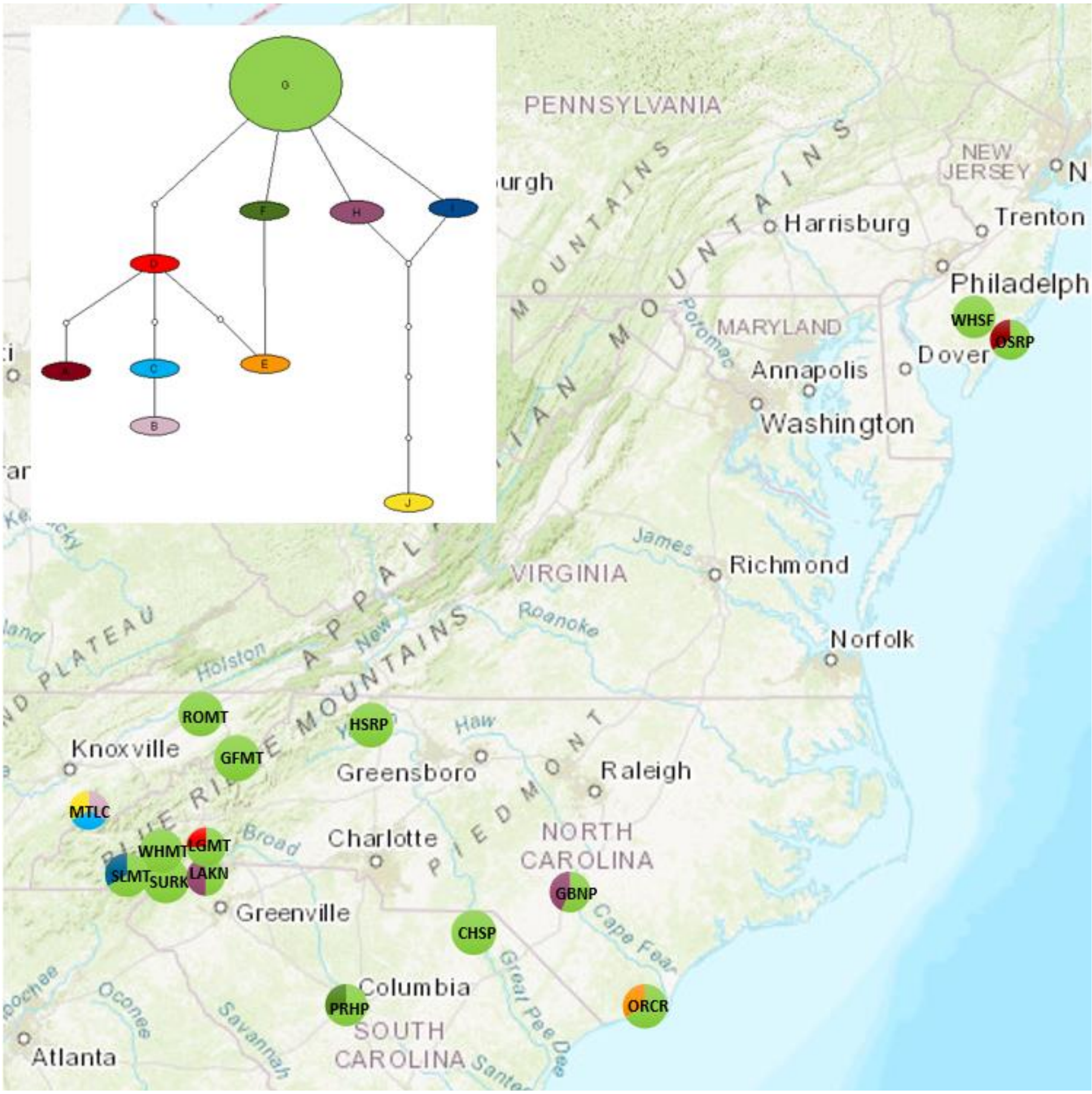


Figure 3. Map of *Kalmia buxifolia* haplotypes generated from sequences of three cpDNA markers (*trnL-F* intron, *trnS-G* intergenic spacer, and *rp*S4 intergenic spacer). Colors represent the 10 groups on the haplotype network (inset) and each pie chart shows the mix of haplotypes found in that population. Each network line segment represents one mutation.

Phylogenetic analysis in Mr. Bayes resulted in nearly identical consensus trees regardless of model implemented (F81 or GTR) and thus trees were combined in Figure 4. Both trees were largely inconclusive with nearly all samples clustering in a polytomy. However, MTLC1 was rooted with the outgroup, *K. procumbens*, rather than grouping with the rest of *K. buxifolia* in both runs (Figure 4). This may support the findings of the haplotype network (Figure 3), where the Mount LeConte haplotype (J) was the most ancestral to *K. procumbens*. Also consistent in both trees, the remaining Mount LeConte samples (MTLC5, MTLC10) group as sister (pp = 0.96, 0.97; Figure 4), which may reflect the uniqueness of the haplotypes in the population found in TCS (Figure 3). The grouping of GBNP1 and LAKN1 as sister to each other (p = 0.77, 0.79; Figure 4) and the larger polytomy also supports the findings of the TCS analysis in which these populations contained the same two haplotypes (G, H), where H is likely a derivative of the most ancestral type (J) found on Mount LeConte.

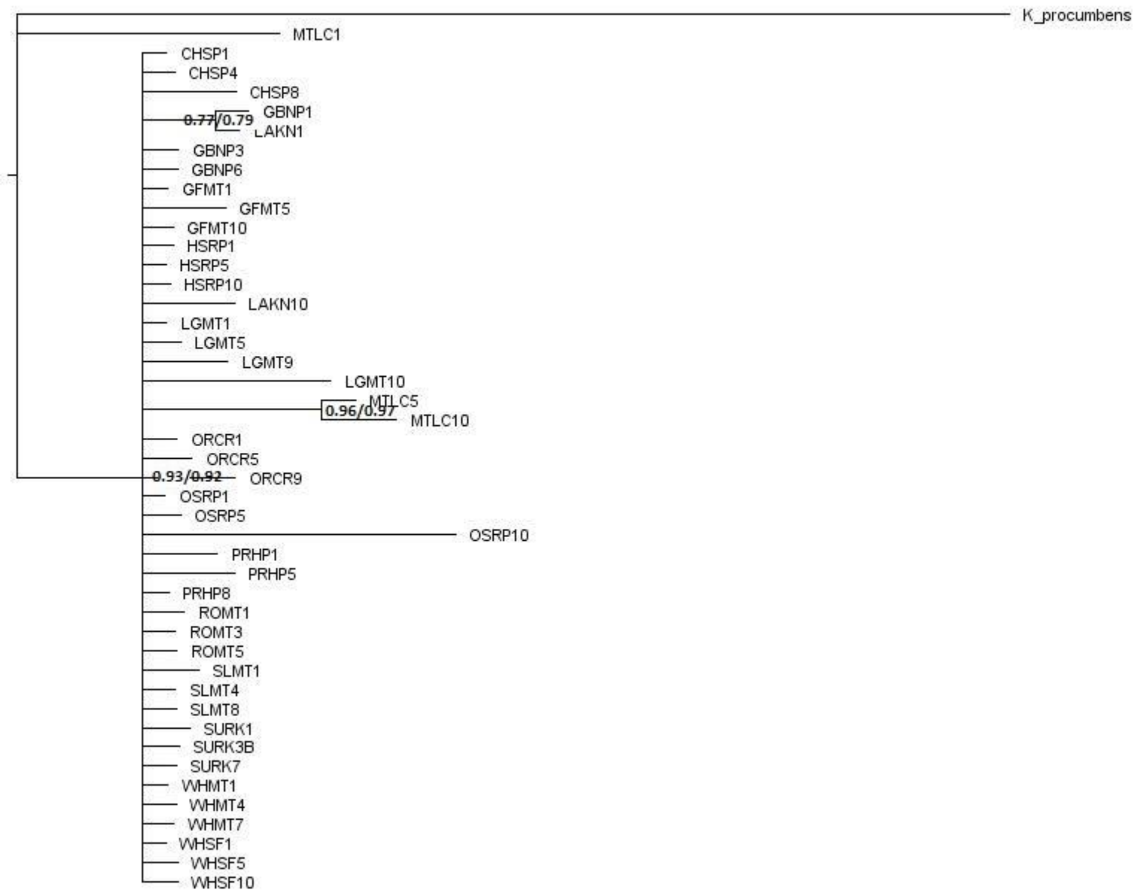


Figure 4. MrBayes consensus tree for *Kalmia buxifolia* based on three cpDNA markers (*trnL-F* intron, *trnS-G* intergenic spacer, and *rp*S4 intergenic spacer) with *K. procumbens* as the outgroup. Posterior probabilities displayed on nodes show results from F81 and GTR models respectively.

Stable Isotope and Stomatal Analysis

A PCA was conducted to examine the relationships between variation among populations in CID and environmental variables (Figure 5). PC axes 1 and 2 accounted for more than 65% of sampling variance (Spring: PC1=41.497%, PC2=23.592, PC3=23.031%, PC4=11.88%; Fall: PC1=41.767%, PC2=23.842%, PC3=21.200%, PC4=13.192%), and held eigenvalues close to 1 in both sampling sets (Figure 2; Spring: PC1=1.660, PC2=0.944, PC3=0.921, PC4=0.475; Fall: PC1=1.671, PC2=0.954, PC3=0.848, PC4=0.528). CID and average soil depth were strongly and positively correlated with PC1, suggesting sites to the right of the axis (i.e. Piedmont and Coastal populations) are characterized by lower WUE (as it is negatively correlated with CID) and deeper soils (Tables 4,5). Similarly, average moss mat depth and average health score were strongly correlated with PC2, although the relationship was positively correlated with the Spring sampling and negatively correlated with the Fall due to the differently rotated axes. Regardless, these relationships followed a similar pattern in both samplings. Samples clustered by region along both axes, with mountain samples exhibiting greater variation (particularly along PC2) than their Coastal counterparts (Figure 5). In both PCAs, some samples from the Mountain set and Piedmont/Coastal set overlapped around 0 along PC1. However, this relationship was not indicative of any population but rather individuals from several populations, demonstrating the variation within both sites and regions rather than populations as a whole.

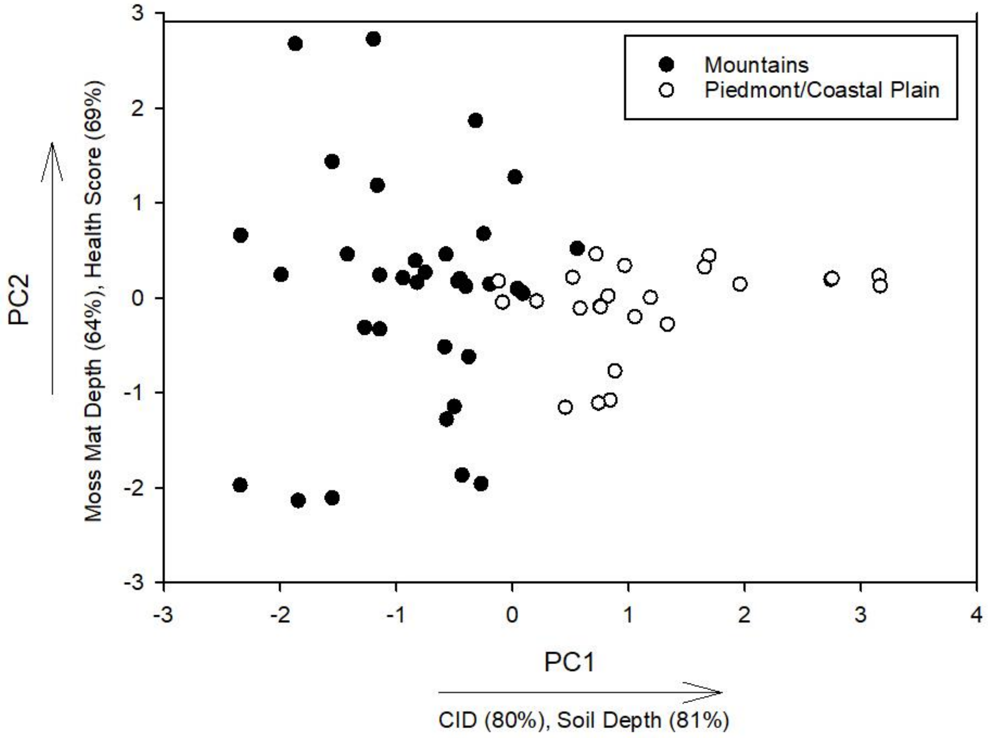
Mean stomatal density of the Mountain and Coastal/Piedmont populations were 53.06 ±14.39 (standard deviation) and 49.22 ±11.61, respectively (Figure 6). By ANOVA, this difference in mean density by region was statistically significant ($p<0.0001$). Additionally, analysis of leaf area found the mean leaf area of Mountain populations to be $0.66cm^2$ and $0.87cm^2$ in Coastal/Piedmont populations. A second ANOVA found these mean differences in leaf area also were significant ($p<0.03$). Finally, regression analysis between CID and stomatal density indicated a significant ($p= 0.0008$), but weak relationship with $r^2 = 0.03$ (Figure 7).

Table 4. Eigenvectors scaled to standard deviation, *Kalmia buxifolia* Spring sampling.

Species	1	2	3	4
Δ	0.7987	-0.1854	0.3291	0.4684
Avg. soil depth	0.8108	0.1445	0.3056	-0.4778
Avg. moss mat depth	-0.4211	0.6388	0.6357	0.1025
Avg. health score	0.4327	0.6931	-0.5616	0.1305

Table 5. Eigenvectors scaled to standard deviation, *Kalmia buxifolia* Fall sampling.

Species	1	2	3	4
Δ	0.7946	0.1544	0.2807	0.5158
Avg. soil depth	0.7430	-0.2606	0.4125	-0.4581
Avg. moss mat depth	-0.4994	-0.7175	0.4358	0.2142
Avg. health score	0.4877	-0.5892	-0.6396	0.0769



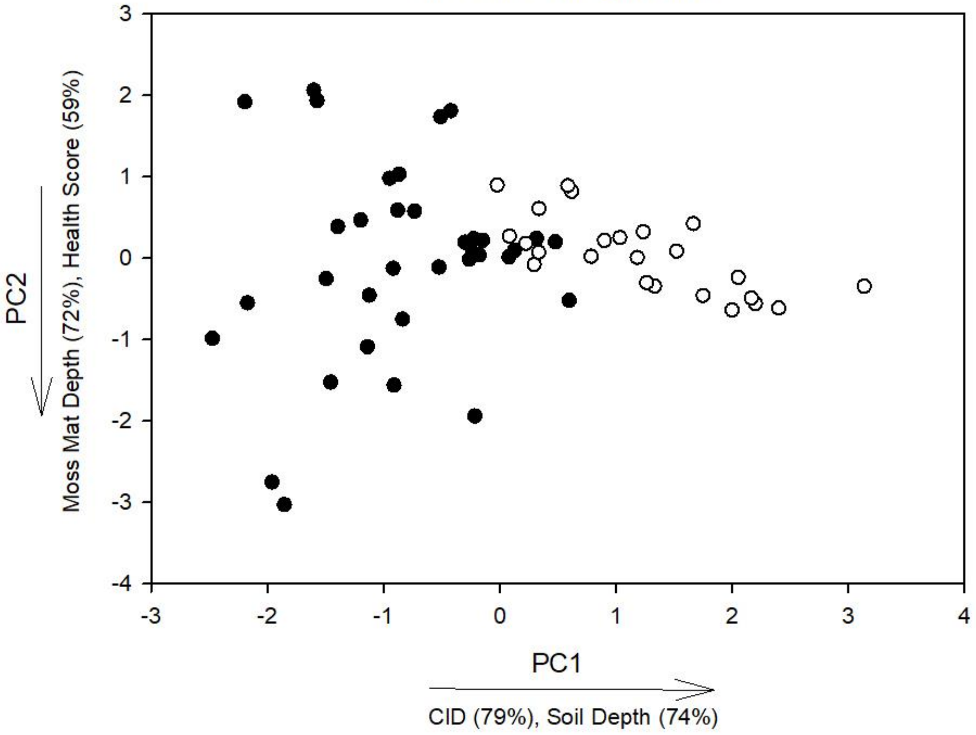


Figure 5. PCA plot of Spring (top) and Fall (bottom) *Kalmia buxifolia* samplings, PC1 v. PC2 with eigenvectors.

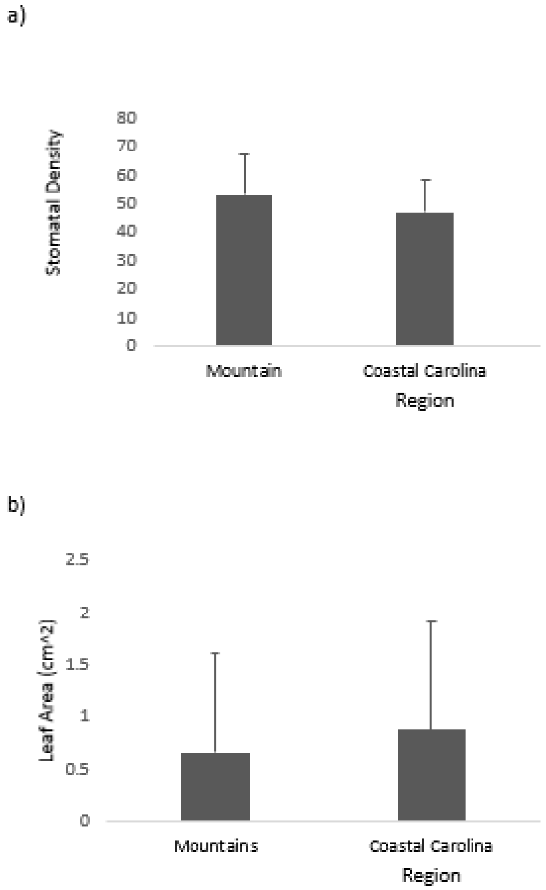


Figure 6. Mean differences in stomatal density (a) and leaf area (b) in *Kalmia buxifolia* by region.

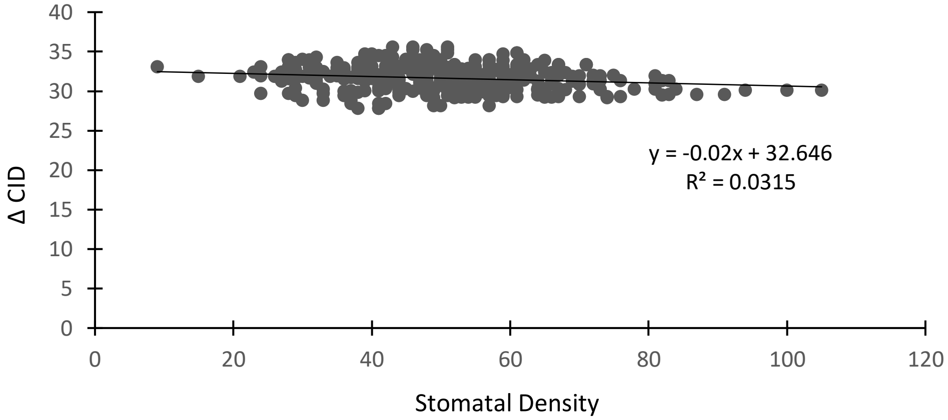


Figure 7. Regression of ΔCID and stomatal density in *Kalmia buxifolia*. Relationship is significant ($p=0.0008$) but weak ($r^2=0.03$). WUE inferred as the inverse relationship of ΔCID.

CHAPTER FOUR: DISCUSSION

Dating analysis of the divergence of *K. buxifolia* and *K. procumbens* refutes the hypothesis that southern Appalachian rock-outcrop species are of Pleistocene origin (maximum 2.58 Ma). Here, the divergence of these sister species is placed much earlier, in the mid-Miocene (~14.9 Ma). Although the 95% HPD (~3.5-28Ma) around this date is wide (due to the wide date placed on the calibrating fossil), it excludes the Pleistocene and late-Pliocene. Such a Miocene origin of *K. buxifolia* is supported by dating of another common rock-outcrop species, *Micranthes petiolaris* (Raf.) Brouillet & Gornall, whose divergence was also placed at c. 14 Ma (Mathews, K., unpublished data). Following paleoclimatic evidence, it seems likely a mid-Miocene divergence for these species was largely brought on by climatic shifts. The period following the Terminal Eocene Event (35 Ma) until the late-Miocene was characterized by frequent oscillations between warm/humid and cool/dry climates that would have facilitated extensive range expansion and contractions (Zachos et al. 2008, Meseguer et al. 2015). The Middle-Miocene Climate Transition (~14 Ma), which was a sudden decrease in average global temperature, is thought to have triggered diversification events across the northern hemisphere (Herbert et al. 2016). Both Wen et al. (2010) and Manos and Meireles (2015) found that the majority of divergences dated for North American – East Asian disjunct clades fall within the Miocene and attributed the splitting largely to climatic oscillations. Manos and Meireles (2015) in particular investigated the origins of over 250 southern Appalachian woody species representing 158 clades and found divergence times for over 60% of the clades placed in the last 20 Myr.

Population-level divergence for *K. buxifolia* was dated to mid-to late-Pliocene, a period also strongly characterized by fluctuations in climate. A major warming period (the Mid-Pliocene Warming Event) occurred between 3-4 Ma, at or slightly before several of the splitting events observed in our analysis. This warming not only resulted in habitat shifts but also a significant rise in sea-level, which is thought to have peaked 30m higher than it is today and likely contributed to extinction and isolation of some of the more coastal populations (Dwyer and Chandler 2009). Late Pliocene – early Pleistocene diversification has been documented in *Sabatia*, another plant group with a similar Appalachian – Coastal Plain disjunction, with the divergence between *S. capitata* (Appalachian species) and *S. gentianoides* (coastal species) occurring around 2.96 mya (Mathews et al. 2015). The end of the Pliocene and beginning of the Pleistocene marked the transition from a warm, tropical climate to a much colder global climate than seen today in the northern hemisphere. The final population divergences in our analysis coincide with this transition, occurring at the start of the Pleistocene (~2 – 1.4 Ma), and were yet again likely in response to changes in the global climate and subsequent habitat range shifts.

Haplotype network analysis further explored the phylogeographic patterns at the population level. The Mount LeConte population in GSMNP not only resolved as the oldest population (~3.38 Ma) of those assessed in this study, but also contained the greatest haplotype diversity (including the most ancestral and most derived types). Thus, in this analysis, Mount LeConte appears to be the oldest remaining refugium and the most historically isolated population. Interestingly, Mount LeConte also is the only population sampled from the western side of the Blue Ridge Divide, which may have contributed to its isolation and haplotype divergence. The Appalachian Mountains are known to have served as a barrier to dispersal for at least the last 300 myrs, and the Mount LeConte population may not have been connected to the other populations since it diverged over 3 Ma. Meanwhile, although the populations on the eastern side of the mountains are dispersed over a wide geographic area, there are few barriers to their dispersal and it is plausible that they may have formed more interconnected populations several times over the last millennia.

Evidence for interconnected populations can also be seen in the haplotype analysis, as all populations (excluding Mount LeConte) contained the same “G” haplotype, which is neither very ancestral or recently derived. Additionally, a Mountain population (Laurel Knob) and a Coastal Plain population (Gordon Butler Nature Preserve) share a unique, ancestral haplotype (“H”), suggesting that at least those two populations were once interconnected. Although Mount LeConte appears to be the most ancestral population sampled, there is also support for at least two more recent Pleistocene refugia in the mountains and along the Sufolk escarpment. Satulah Mountain, Laurel Knob, and Little Green Mountain all exhibited some haplotype diversity, with the common "G" haplotype and a second unique type. Of these, the unique types found on Satulah ("I") and Laurel Knob ("H") are both derivatives of the most ancestral type found on Mount LeConte ("J"), while Little Green ("D") is a precursor to the derived types on Mount LeConte ("B" and "C"). Thus, these mountain sites (particularly Satulah and Laurel Knob) may be representative of a second refugium retaining these older haplotypes or possibly evidence for greater gene flow between the western and eastern sides of the mountains than expected. The third potential refugium lies along the escarpment, with Peachtree Rock Heritage Preserve ("F") and Gordon Butler Nature Preserve ("H") also exhibiting a second haplotype. Coastal flooding over the Pliocene and Pleistocene would have wiped out any populations that established east of this line, but the Sufolk escarpment itself was never underwater. Therefore, these populations may be remnants of a once more robust coastal population. The most coastal population, Orton Creek, also exhibits a unique haplotype ("E"), however this type is a derivative of the Peachtree Rock type and likely emerged after sea-level subsided and the coast was recolonized from the remaining escarpment populations.

The two New Jersey populations sampled (Wharton State Forest and Oswego River Preserve) contained the common "G" haplotype as well as second type ("A") found only in the Oswego River population. This "A" haplotype is one of the most derived in the network and a derivative of the "D" type found in the Little Green Mountain population. This supports the hypothesis that the Pine Barrens populations were colonized via a mountain corridor, potentially by dispersal from the historical populations in Pennsylvania. Additionally, BEAST analysis placed the divergence of these populations in the early Pleistocene along with the southern populations, and the Pine Barrens are known to never have flooded (being structurally high like the Sufolk escarpment; Stanford 2015) and although periglacial, were never under ice. It is possible that these sites were colonized via a mountain corridor during the late-Pliocene, early-Pleistocene, became isolated during Pleistocene glaciation, and diverged.

Results of the CID analysis suggest a long isolation of regional *K. buxifolia* populations at least across North Carolina, which has led to differential selection in ecophysiology and morphological traits related to WUE. However, over the 2017 season, Piedmont and Coastal populations had lower WUE (higher CID) than their Mountain counterparts; this trend is counter to the hypothesis that Mountain populations would have lower WUE due to daily fog immersion, which can ameliorate drought stress in plants through both attenuating direct insolation and as a source of moisture via foliar-uptake. Further, analyses revealed a weak, negative relationship between stomatal density and CID, and greater mean stomatal density in Mountain populations; these results are counter to an expectation of lower stomatal density (or fewer stomata) in more drought-stressed plants (Milburn and Weatherley 1971). A greater standard deviation of stomatal density among Mountain populations suggests greater variation among individuals than their Coastal counterparts and may be due to the high variability in microclimates found in mountain systems. The Coastal samples however, had a larger mean leaf area, which is common in shade-plants and may reflect their more closed Pine-Savanna habitats, in contrast to exposed high-elevation rock outcrops.

Still, the question of the geographical pattern of WUE in *K. buxifolia* remains. One hypothesis is that without cloud immersion, Coastal plants may close their stomates early each day to cope with midday heat and water stress (e.g. *Allium cepa*, *Quercus suber*, etc.) (Meidner and Heath 1959, Tenhunen et al. 1983), thus reducing CO_2 uptake and daily net photosynthesis. Meanwhile, Mountain plants may leave their stomata open longer, especially on cloudy or foggy days. The generally high humidity in the Mountain region may also reduce the leaf-air vapor pressure gradient and water loss across stomates while maintaining CO_2 uptake and photosynthesis, thus producing greater daily WUE (net photosynthesis : transpiration). Further stomatal analysis revealed that the average number of stomata was not statistically different between regions, supporting the hypothesis that it is stomatal activity rather than morphology that is important for regulating water loss and WUE (one-way ANOVA, p=0.64). Horton and Culatta (2016) showed that on cloudy days, outcrop specialists (*Hydatica petiolaris* and *Sibbaldiopsis tridentata*) had a lower leaf-to-air vapor pressure deficit, higher stomatal conductance, and higher internal CO_2 concentration. However, WUE was not consistently higher for the outcrop specialists, although it varied by species and may be higher *in K. buxifolia*. If this hypothetical trend is true and continues, Coastal plants may open their stomata for shorter and shorter periods as the climate warms, and eventually might reach a point where they are no longer open long enough to maintain carbon gain. Additionally, if fog disappears from the mountains, plants will become exposed to greater insolation and may begin regulating water loss like the Coastal plants with mid-day stomatal closures (Richardson et al. 2003, McNulty et al. 2012). To test this hypothesis, much more work is needed to understand both conductance and photosynthetic variation in these plants, including developing daily photosynthetic and transpiration curves over seasonal variation in temperature and soil moisture. Nevertheless, current variation in WUE suggests temperature and soil moisture patterns since population divergence in the Pleistocene have led to differences in ecophysiological and morphological traits related to water use.

CHAPTER FIVE: CONCLUSIONS

The objectives of this investigation were twofold. First, I aimed to reconstruct the phylogeographic history of *K. buxifolia* through determining locations of glacial refugia, directions of postglacial colonization, and dating the divergence from the sister species, *K. procumbens*. The divergence was dated to 14.9 Ma (95% HPD: 3.5 – 28 Ma), placing it likely mid-Miocene and much later than predicted by the rock-outcrop hypothesis. This divergence, as well as population level divergences in this species, seems to coincide with major climatic shifts from the mid-Miocene to early-Pleistocene. Four potential locations of glacial refugia were identified, with the most ancient on the western side of the mountains at Mount LeConte, and more recent refugia on the eastern side of the mountains surrounding the Highlands Plateau (Satulah Mountain, Laurel Knob, and Little Green Mountain), along the Sufolk escarpment (Peachtree Rock Heritage Preserve and Gordon Butler Nature Preserve), and in the New Jersey Pine Barrens (Oswego River Preserve). Directions of postglacial colonization are difficult to infer due to relatively little within and among population variation in this species. However, the haplotype and dating analysis does suggest the Pine Barrens populations are of mountain origin, becoming isolated during the Pleistocene.

To further examine these population level divergences, the second objective was to assess whether isolation has led to differing environmental selection, particularly in traits related to water-use. Thus, variation in stress tolerance through water-use efficiency (WUE) was assessed for the mountain and coastal populations of *K. buxifolia* in the Carolinas over a season. Here, stable isotope analysis revealed geographical variation in WUE, although contrary to our hypothesis, as the Mountain populations exhibited greater WUE than the coast. Further research is necessary to understand how and why this variation occurs, particularly in the face of a once-again changing climate. Such information may be important for developing management plans, not only for *K. buxifolia* but rock-outcrop communities as a whole, and for making predictions on the future of the species.

LITERATURE CITED

Akaike, H. (1974). A New Look at the Statistical Model Identification. *Springer Series in Statistics Selected Papers of Hirotugu Akaike*, 215-222. doi:10.1007/978-1-4612-1694-0_16

Bergius, P. J. (1777). *Ledum buxifolium*. *Acta Acad. Sci. Imp. Petrop,* 1(1): 213.

Berry, Z. C. & W. K. Smith. (2012). Cloud pattern and water relations in *Picea rubens* and *Abies fraseri*, southern Appalachian Mountains, USA. *Agricultural and Forest Meteorology*, 162-163:27–34.

Berry, Z. C. & W. K. Smith. (2014). Experimental cloud immersion and foliar water uptake in saplings of Abies fraseri and Picea rubens. *Trees*, 28(1), 115-123. doi:10.1007/s00468-013-0934-5

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C-H., Xie, D., Suchard, M. A., Rambaut, A., & A. J. Drummond. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology,* 10(4), e1003537. doi:10.1371/journal.pcbi.1003537

Camp, WH. (1938). The genus *Leiophyllum*. *Bull. Torrey Bot. Club,* 65:99-104.

Campbell, J., & M. Medley. (2012). The atlas of vascular plants in Kentucky.

Chernykh, I. V., Alduchov, O. A., & R. E. Eskridge. (2001). Trends in low and high cloud boundaries and errors in height determination of cloud boundaries. *Bull. Amer. Meteor. Soc*, (82), 1941-1947.

Clement M., Posada D. & K. Crandall. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, 9(10): 1657-1660

Culatta K.E. & J. L. Horton. (2014). Physiological Response of Southern Appalachian High-Elevation Rock Outcrop Herbs to Reduced Cloud Immersion. *Castanea*, 79(3):182–194.

Divins, D. L. & D. Metzger. (2008). Coastal relief model: National Geophysical Data Center, Available at: http://www.ngdc.noaa.gov/mgg/coastal/coastal.html.

Dorken, M. E., & S. C. Barrett. (2004). Chloroplast haplotype variation among monoecious and dioecious populations of *Sagittaria latifolia* (Alismataceae) in eastern North America. *Molecular Ecology*,13(9), 2699-2707. doi:10.1111/j.1365-294x.2004.02246.x

Dwyer, G. S. & M. A. Chandler. (2009). Mid-Pliocene sea level and continental ice volume based on coupled benthic Mg/Ca palaeotemperatures and oxygen isotopes. *Philosophical Transactions of the Royal Society A*, (367) 157-168.

Elliott, S. (1817). A sketch of the botany of South Carolina and Georgia. J. R. Schenck, Charleston, S. C.

Ferguson, A. (1980). Biochemical systematics and evolution. Glasgow: Blackie.

Foster, P. (2001). The potential negative impacts of global climate change on tropical montane cloud forests. *Earth-Science Reviews*, (5)73-106.

French, H. M., Demitroff, M., & S. L. Forman. (2003). Evidence for late-Pleistocene permafrost in the New Jersey Pine Barrens(latitude 39°N), eastern USA. *Permafrost and Periglacial Processes*,14(3), 259-274. doi:10.1002/ppp.456

Gillespie, E. L. & K.A. Kron. (2013). Molecular Phylogenetic Relationships and Morphological Evolution within the Tribe Phyllodoceae (Ericoideae, Ericaceae). *Systematic Botany*, 38(3), 752-763.

Godt, M. J., Johnson, B. R., & J. Hamrick. (1996). Genetic Diversity and Population Size in Four Rare Southern Appalachian Plant Species. *Conservation Biology*,10(3), 796-805. doi:10.1046/j.1523-1739.1996.10030796.x

Herbert, T. D., Lawrence, K. T., Tzanova, A., Peterson, L. C., Caballero-Gill, R., & C. S. Kelly. (2016). Late Miocene global cooling and the rise of modern ecosystems. *Nature Geoscience*,9(11), 843-847. doi:10.1038/ngeo2813

Hewitt, G. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*,58(3), 247-276. doi:10.1006/bijl.1996.0035

Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*,68(1-2), 87-112. doi:10.1111/j.1095-8312.1999.tb01160.x

Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), 907-913. doi:10.1038/35016000

Ho, S. Y. W. (2007). Calibrating molecular estimates of divergence times in birds. *Journal of Avian Biology*, 38: 409-414.

Horton J. L. & K. E. Culatta. (2016). Physiological Characteristics of Southern Appalachian High-Elevation Rock Outcrop Herbs on Clear and Cloudy Days. *Castanea*, 81(4): 270-279

Huelsenbeck, J. P. & F. Ronquist. (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics,* 17:754-755.

Ikeda, H., Eidesen, P. B., Yakubov, V., Barkalov, V., Brochmann, C., & H. Setoguchi. (2017). Late Pleistocene origin of the entire circumarctic range of the arctic-alpine plant *Kalmia procumbens*. *Molecular Ecology*, 26(20), 5773-5783. doi:10.1111/mec.14325

Ikeda, H., & H. Setoguchi. (2006). ORIGINAL ARTICLE: Phylogeography and refugia of the Japanese endemic alpine plant, *Phyllodoce nipponica* Makino (Ericaceae). *Journal of Biogeography*, 34(1), 169-176. doi:10.1111/j.1365-2699.2006.01577.x

Kartesz, J.T. (2014). The Biota of North America Program (BONAP). Taxonomic Data Center. (http://www.bonap.net/tdc). Chapel Hill, N.C. [maps generated from Kartesz, J.T. 2015. Floristic Synthesis of North America, Version 1.0. *Biota of North America Program* (BONAP).

Kron, K. A., & J. M. King. (1996). Cladistic Relationships of *Kalmia, Leiophyllum*, and *Loiseleuria* (Phyllodoceae, Ericaceae) Based on rbcL and nrITS Data. *Systematic Botany*, 21(1), 17. doi:10.2307/2419560

King, R. A., & C. Ferris. (1998). Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology*, 7(9), 1151-1161. doi:10.1046/j.1365-294x.1998.00432.x

Limm, E. B., & T. E. Dawson. (2010). *Polystichum munitum* (Dryopteridaceae) varies geographically in its capacity to absorb fog water by foliar uptake within the redwood forest ecosystem. *American Journal of Botany*,97(7), 1121-1128. doi:10.3732/ajb.1000081

Mai, D.H. (2001). Die mittelmiozänen und obermiozänen Floren aus der Meuroer und Raunoer Folge in der Lausitz. Teil II: Dicotyledonen. *Palaeontographica Abterilung B*, (257) 35-174.

Manos, P.S. & J.E. Meireles. (2015). Biogeographic analysis of the woody plants of the southern Appalachians: implications for the origins of a regional flora. *American Journal of Botany*, 102(5), 1-804).

Mathews, K. G., Ruigrok, M. S., & G. Mansion. (2015). Phylogeny and Biogeography of the Eastern North American Rose Gentians (Sabatia, Gentianaceae). *Systematic Botany*, 40(3), 811-825. doi:10.1600/036364415x689276

McCune, B. & M. J. Mefford. (2016). PC-ORD. Multivariate Analysis of Ecological Data. Version 7. MjM Software Design, Gleneden Beach, Oregon, U.S.A.

McNulty, S., Myers, J. M., Caldwell, P., & G. Sun. (2012). Climate change, Chapter 3*, in* Southern Forest Futures Project Technical Report. Available online http://www.srs.fs.usda.gov/futures/reports/draft/Frame.htm.

Medley, M.E. (1993). An annotated catalogue of the known or reported vascular flora of Kentucky. University of Louisville, Louisville.

Meidner, H. & O.V.S. Heath. (1959). Studies in stomatal behavior: VIII. Stomatal responses to temperature and carbon dioxide concentration in *Allium cepa* L., and their relevance to mid-day closure. *Journal of Experimental Botany*, 10(2), 206-219.

Meseguer, A. S., Lobo, J. M., Ree, R., Beerling, D. J., & Sanmartín, I. (2015). Integrating Fossils, Phylogenies, and Niche Models into Biogeography to Reveal Ancient Evolutionary History: The Case of *Hypericum* (Hypericaceae). *Systematic Biology*, 64(2), 215-232. doi:10.1093/sysbio/syu088

Michener R. & K. Lajtha. (2007, edited). *Stable Isotopes in Ecology and Environmental Science*. Blackwell Publishing Ltd., 2007.

Milburn, J. A., & P. E. Weatherley. (1971). The Influence Of Temperature On The Process Of Water Uptake By Detached Leaves And Leaf Discs. *New Phytologist*, 70(5), 929-938.

Miller, M. A., Pfeiffer, W., & T. Schwartz. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA pp 1 - 8.

Parham, P. R., Riggs, S. R., Culver, S. J., Mallinson, D. J., Rink, W. J., & Burdette, K. (2012). Quaternary coastal lithofacies, sequence development and stratigraphy in a passive margin setting, North Carolina and Virginia, USA. *Sedimentology*, 60(2), 503-547. doi:10.1111/j.1365-3091.2012.01349.x

Persoon, C. H. (1805). *Syn. Pl.*, 1: 147. C. F. Cramerum, Paris

Petit, R. J. (2003). Glacial Refugia: Hotspots But Not Melting Pots of Genetic Diversity. *Science*, 300(5625), 1563-1565. doi:10.1126/science.1083264

Petit, R. J., Brewer, S., Bordács, S., Burg, K., Cheddadi, R., Coart, E., . . .& A. Kremer. (2002). Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management*, 156(1-3), 49-74. doi:10.1016/s0378-1127(01)00634-x

Potzger, J. (1945). The Pine Barrens of New Jersey, A Refugium During Pleistocene Times. Butler University Botanical Studies, 7(1/13), 182-196. Retrieved from http://www.jstor.org/stable/41822412

Provan, J., & K. Bennett (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, 23(10), 564-571. doi:10.1016/j.tree.2008.06.010

Pisias, N., & T. Moore. (1981). The evolution of Pleistocene climate: A time series approach. *Earth and Planetary Science Letters*, 52(2), 450-458. doi:10.1016/0012-821x(81)90197-7

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & M. A. Suchard. (2013) Posterior summarisation in Bayesian phylogenetics using Tracer 1.6. *Systematic Biology*. syy032. doi:10.1093/sysbio/syy032

Rhoads, A.F., & W.M. Klein, Jr. (1993). *The vascular flora of Pennsylvania: Annotated checklist and atlas.* American Philosophical Society, Philadelphia.

Richardson, A. D., Denny, E. G., Siccama, T. G., Lee, X. (2003). Evidence for a Rising Cloud Ceiling in Eastern North America**. Journal of Climate*, 16(12):2093–2098.

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., Siripun, K. C., Winder, C. T., Schilling, E., & R. L. Small. (2005). The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, 92(1): 142-166.

Shaw, J., Lickey, E. B., Schilling, E. E., & R. L. Small. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in Angiosperms: The tortoise and the hare III. *American Journal of Botany*, 94(3): 275-288.

Small, J. K. (1903*). Flora of the southeastern United States*. New York: Published by Author.

Smith, B. T., Escalante, P., Baños, B. E., Navarro-Sigüenza, A. G., Rohwer, S., & Klicka, J. (2011). The role of historical and contemporary processes on phylogeographic structure and genetic diversity in the Northern Cardinal, Cardinalis cardinalis. *BMC Evolutionary Biology*, 11(1). doi:10.1186/1471-2148-11-136

Soltis, D. E., Morris, A. B., Mclachlan, J. S., Manos, P. S., & P. S. Soltis. (2006). Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, 15(14), 4261-4293. doi:10.1111/j.1365-294x.2006.03061.x

Stanford, S. D. (2015). Groundwater seepage, landforms, and landscape evolution in the New Jersey Pine Barrens. Abstracts With Programs - *Geological Society Of America*, 47(7), 821.

Stehlik, I., Schneller, J. J., & K. Bachmann. (2001). Resistance or emigration: response of the high-alpine plant *Eritrichium nanum* (L.) Gaudin to the ice age within the Central Alps. *Molecular Ecology*, 10(2), 357-370. doi:10.1046/j.1365-294x.2001.01179.x

Strand A. E. & R. Wyatt. (1991). Geographical Variation and Biosystematics of Sand Myrtle, *Leiophyllum buxifolium* (Ericaceae). *Systematic Botany*, 16(3):529.

Sun, B., Karl, T. R., & D. J. Seidel. (2007). Changes in Cloud-Ceiling Heights and Frequencies over the United States since the Early 1950s. *Journal of Climate*, 20(15), 3956-3970. doi:10.1175/jcli4213.1

Taberlet, P., Gielly, L., Pautou, G., & J. Bouvet. (1991). Universal primers for amplification of three noncoding regions of chlororplast DNA. *Plant Molecular Biology*, 17: 1105-1109.

Taberlet, P., Fumagalli, L., Wust-Saucy, A., & J. Cosson. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, 7(4), 453-464. doi:10.1046/j.1365-294x.1998.00289.x

Tenhunen, J.D., Lange, O. L., Gebel, J., Beyschlag, W., & J.A. Weber. (1984). Changes in photosynthetic capacity, carboxylation efficiency, and CO_2 compensation point associated with midday stomatal closure and midday depression of net CO_2 exchange of leaves of *Quercus suber*. *Planta*, 162(3), 193-203.

Thompson, S.A. (1997). *Vascular plants: review of status in Pennsylvania*. Department of Conservation and Natural Resources, Pennsylvania.

US Geological Survey. (2012, May 30). Glaciers and Sea Level. Retrieved from http://www.usgs.gov/climate_landuse/glaciers/glaciers_sea_level.asp

Vendramin, G. G., Degen, B., Petit, R. J., Anzidei, M., Madaghiele, A., & B. Ziegenhagen. (1999). High level of variation at *Abies alba* chloroplast microsatellite loci in Europe. *Molecular Ecology*, 8(7), 1117-1126. doi:10.1046/j.1365-294x.1999.00666.x

Wear, D. N. & J.G. Greis. (2012). The Southern Forest Futures Project: summary report. Gen. Tech. Rep. SRS-GTR-168. Asheville, NC: USDA-Forest Service, Southern Research Station. 54 p.

Wen, J., Ickert-Bond, S., Nie, Z., & R. Li. (2010). Timing and modes of evolution of Easter Asian – North American biogeographic disjunctions in seed plants. (*In* M. Long, H. Gu, and Z. Zhou [eds].), *Darwin's heritage today: Proceedings of the Darwin 200 Beijing International Conference*, 252-269. Higher Education Press, Beijing, China.

White, T., Bruns, T., Lee, S., & J. Taylor. (1990). Amplification And Direct Sequencing Of Fungal Ribosomal Rna Genes For Phylogenetics. *PCR Protocols*, 315-322. doi:10.1016/b978-0-12-372180-8.50042-1

Wilbur, R. L. & C. H. Racine. (1971). The genus Leiophyllum (Ericaceae): Morphological variation, distribution, and classification. *J. Elisha Mitchell Sci. Soc*. (87), 222-226.

Wiser, S. K. (1994). High-elevation cliffs and outcrops of the southern Appalachians: Vascular plants and biogeography. *Castanea*, (59), 85-116.

Wiser, S. K. (1998). Comparison of Southern Appalachian high-elevation outcrop plant communities with their Northern Appalachian counterparts. *Journal of Biogeography*, 25(3), 501-513. doi:10.1046/j.1365-2699.1998.2530501.x

Wiser S. K., Peet R. K., & P. S. White. (1996). High-elevation rock outcrop vegetation of the Southern Appalachian Mountains. *Journal of Vegetation Science*, 7(5):703–722.

Wood, C. E. (1961). The genera of Ericaceae in the southeastern United States. *J. Arnold Arbor.,* (42), 10-73

Yang, Z., & B. Rannala. (1997). Bayesian phylogenetic inference using DNA sequences: A Markov Chain Monte Carlo Method. *Molecular Biology and Evolution*, 14(7), 717-724. doi:10.1093/oxfordjournals.molbev.a025811

Zachos, J. C., Dickens, G. R., & R. E. Zeebe. (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature*, 451(7176), 279-283. doi:10.1038/nature06588

APPENDIX

Nexus File of Aligned Sequences

```
#NEXUS

[3 cpDNA regions - trnLF: 1-520; trnSG: 521-1398; rpS4: 1399-1938]

begin taxa;
    dimensions ntax=46;
    taxlabels
K_procumbens
CHSP1
CHSP4
CHSP8
GBNP1
GBNP3
GBNP6
GFMT1
GFMT5
GFMT10
HSRP1
HSRP5
HSRP10
LAKN1
LAKN10
LGMT1
LGMT5
LGMT9
LGMT10
MTLC1
MTLC5
MTLC10
ORCR1
ORCR5
ORCR9
OSRP1
OSRP5
OSRP10
PRHP1
PRHP5
PRHP8
ROMT1
ROMT3
ROMT5
SLMT1
SLMT4
SLMT8
SURK1
SURK3B
SURK7
WHMT1
WHMT4
WHMT7
WHSF1
WHSF5
WHSF10
;
```

```
end;

begin characters;
  dimensions nchar=1938;
  format datatype=dna gap=- interleave=no;
  matrix

K_procumbens
TCAGAGAAACCCCGGAAAAAAAAAAATGGGCAATCCTGAAGCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGCAAAAAAAAAAAAG
GATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGGTTGGTAGAATAATAGAAATCCTTTTCTCGAAACTTACGAAAA
AATGAAA-GATAACTCT--AT---T-C-TATA--TA--TATATATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCTTTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTGAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAA-----------
GATAATTACTGTATTACATTACACGGAAAGTAAGGATTCCAAAAATTCTTTCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA-
--T-----A--AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAAAA-AA-
AATTCTTTGCTTTCTTTTTGTACA-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCAAAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAAAAAAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAAACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCGAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAA-
AAACCCCCTTATTTAATTTGAATTGATCCCTCTTTTCGAAACCTCCTTAACCTCGCACAAAAAAANGGGAACACTC-
ANTTTTCCAGGATTTCTTTT-AGGAT-CTT-AT-C-TTGATANCN-CCCNNCTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATA-
TATATTACATTGTAGCGG-G--TATACTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAGAGGGTCTTTCGGTTTAATTCATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAACTCGATTCCGATAGCCGGCTTTTCTCTATTTCGATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGA
ATATTGAAAAAACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTATT
TTTTTTCTAGAATATAATAGAAAGGGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCAT
TTAGTTCAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTC-
AAAAAAAAAATTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTTATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGC---
TTTACCAGGACTAACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCAATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGT
CTACTAGAAAAACAAAAATTGCGGTTTCATTACGGACT
CHSP1                   TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--RAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCAWARATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAARAARAAAGTTKGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAAKGGAATAATAACKGGAATAAAAAAGGAAACTATGGATTATTARACAATGCATTTTTATGTTAT
GATTTCGGKGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAARAAAAAAAAAR-AA-C-
CCCTTATTTAAATTGAAAKGAATCCYCTTTTCCRAAYCYCATTAAMCYCCCMCWAAAAAA-GKGAACMCYCRRKTTTTC-AKGATT-CTTTT-
AGGAT-CCT-AT-C-TKGATTACS-CCCAACTCCCTKGTTTGACAAAARATCCWKGATT--AT-ATAAWA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTANTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGA-T-
CGAGGAAAAACTCGATTCCGATAGCCGGCTTTTCTCTATTTCGATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGA
ATATTGAAAAAACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTT
TTTTTTCTAGAATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCAT
TTAGTTCAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTCAAAAAAAAAAAATTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTT
ATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTTTACCAGGACTAACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCA
ATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCTACTAGAAAAACAAAAATTGCGGTTTCATTACGGACT
CHSP4                   TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
```

```
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
CHSP8                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAA-------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
-----------------------------------------
GBNP1                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--RAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCAWARATAAAACRATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAARAARAAAGTTKGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAAKGGAATAATAACKGGAATAAAAAAGGAAACTATGGATTATTARACAAKGCATTTTTATGTTAT
GATTTCGGKGTTTTTATCCAACCGTATCTWTCCAAATTCTTCCAGTAAAAGAAAAAAAAAAARAA-C-
CCCTTATTTAAWTTGAAAGGAAYCCYYTTTYCCAAAYCYCWTWAAMCYCCCMCWAAAAAR-KKGAACMCYCRRKTTTTC-WKGATT-CTTTT-
AGGAT-CCT-AT-C-TKGATTACS-CCCAACTCCCTKGTTTGACAAAARATCCWKGATT--AT-AWAAWA-ATTACTTTGWASCGG-G--
```

```
TATAGTTTA-G-GGTAAAAGTGTGATT-CGTTC---TATTAACCCCNTTTAA--
TAGTTAAAGGGTCTTTCGGTTTANTTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACNGAAT-
CGAGGAAAAACTCGATTCCGATAGCCGGCTTTTCTCTATTTCSATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGA
ATATTGAAAAAACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTT
TTTTTTCTAGAATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCAT
TTAGTTCAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTCAAAAAAAAAAATTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTT
ATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTTTACCAGGACTAACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCA
ATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCTACTAGAAAAACAAAAATTGCGGTTTCATTACGGACT
GBNP3                      TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAC------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
GBNP6                      TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAC------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
GFMT1                      TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
```

```
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--RAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCAWARATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAARAARAAAGTTKGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAAKGGAATAATAACKGGAATAAAAAAGGAAACTATGGATTATTARACAAKGCATTTTTATGTTAT
GATTTCGGKGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAARAAAAAAAAAR-AA-C-
CCCTTATTTAAWTTGAAAKGAAYCCYCTTTTCCRAAYCYCWTWAAMCYCCCMCWAAAAAA-GKGAACMCYCRRKTTTTC-AKGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACS-CCCAACTCCCTKGTTTGACAAAARATCCWTGATT--AT-ATAAWA-ATTACATTGTAGCGG-G--
TATAGTTTANNNGGTAAAAGTGTGATT-CGT-CNNNTATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAACTCGATTCCGATAGCCGGCTTTTCTCTATTTCGATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGA
ATATTGAAAAAACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTT
TTTTTTCTAGAATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCAT
TTAGTTCAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTC-AAAAAAAAA-
TTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTTATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTTTACCAGGACT
AACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCAATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCTACTAGAAAAA
CAAAAATTGCGGTTTCATTACGGACT
GFMT5                ----AGAAACCCCGGAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAA---------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
---------------------------------------
GFMT10               TCAGAGAAACCCCGGAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
```

```
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
HSRP1                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
HSRP5                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAACTCGATTCCGATAGCCGGCTTTTCTCTATTTCGATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGA
ATATTGAAAAAACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTT
TTTTTTCTAGAATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCAT
TTAGTTCAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTV-
AAAAAAAAAATTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTTATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTT
TACCAGGACTAACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCAATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCT
ACTAGAAAAACAAAAATTGCGGTTTCATTACGGACT
HSRP10                TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
```

```
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
LAKN1                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--RAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCAWARATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAARAARAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAAKGGAATAATAACKGGAATAAAAAAGGAAACTATGGATTATTARACAAKGCATTTTTATGTTAT
GATTTCGGKGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAARAAAAAAAAAR-AA-C-
CCCTTATTTAAWTTGAAAKGAAYCCYCTTTTCCRAATCTCATTAAACYCCCMCWAAAAAA-GKGAACMCYCRRKTTTTC-AKGATT-CTTTT-
AGGAT-CCT-AT-C-TKGATTACS-CCCAACTCCCTKGTTTGACAAAARATCCWKGATT--AT-ATAAWA-ATTACTTTGTAGCGG-G--
TATAGTTTA-GNGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTTTCTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAACTCGATTCCGATAGCCGGCTTTTCTCTATTTCGATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGA
ATATTGAAAAAACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTT
TTTTTTCTAGAATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCAT
TTAGTTCAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTCNAAAAAAAAAAATTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTT
ATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTTTACCAGGACTAACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCA
ATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCTACTAGAAAAACAAAAATTGCGGTTTCATTACGGACT
LAKN10                -----GAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAA-------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
```

```
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
-----------------------------------------
LGMT1                 TCAGAGAAACCCCGGAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAACTCGATTCCGATAGCCGGCTTTTCTCTATTTCGATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGA
ATATTGAAAAAACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTT
TTTTTTCTAGAATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCAT
TTAGTTCAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTC-
AAAAAAAAAATTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTTATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTT
TACCAGGACTAACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCAATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCT
ACTAGAAAAACAAAAATTGCGGTTTCATTACGGACT
LGMT5                 ----AGAAACCCCGGAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
LGMT9                 ----AGAAACCCCGGAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
```

```
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAA-------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
-----------------------------------------
LGMT10                ------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
------------------------------------------------------------------------
TGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTTTCTTTCTCTCTTTATCT--AT-ATAT-TATA-
TAG-ATA--TA---T-----A--AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--
GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGAAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTATTTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAANGACCGAAT-
CGAGGAAAAAC-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
MTLC1                 TCAGAGAAACCCCGGAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAAAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--TA--TATAGGTAATAAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTGAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAAAARAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCSCCGGBCGGGCTAGGTTAGWAACTAACCGGGCCCTTTTTTTGTTCCAMCKAAYCAWARAWAAAACRATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAARAAAAAAGTTKGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAAWACAAAAKGGAAWAAWAACKGGAAWAAAAAAGGAAACTAKGGATTATTAAACAAGGCATTTTTATGTTAK
GATTTCGG-GGTTTTWYCCAMCCGTAYYTWYCCAAATTCTTCCAGWAAAARAAAAAAAAAR-AM-C-
CCCTTWTTWAAATKGAAAKGAAYCCYYTTTTCCAAAYCYCWTTWAMCYCCCMCWAAAAAA-GGGAAC-CWCRRKTTTTC-DKGATT-CTTTT-
AGGAT-CCT-AT-C-TKGATTA------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
```

```
-----------------------------------------------------------
TCGATTCCGATAGCCGGCTTTTCTCTATTTCSATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGAATATTGAAAAA
ACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTTTTTTT-
CTAGAATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCATTTAGTT
CAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTC-
AAAAAAAAAATTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTTATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTT
TACCAGGACTTACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCAATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCT
ACTAGAAAAACAAAAATTGCGGTTTCATTACGGACT
MTLC5                 TCAGAGAAACCCCGGAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAAAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--TA--TATAGGTAATAAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTGAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAA--------------
AATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTTTCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---
T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAAAAGAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
MTLC10                TCAGAGAAACCCCGGAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAAAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--TA--TATAGGTAATAAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTGAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAAAAGAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGAATAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
ORCR1                 TCAGAGAAACCCCGGAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
```

```
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--TAGTTAAAGGGTCTTT-G-
TTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-CGAGGAAAAAC----------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
------------------------------------------------------------
ORCR5                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAAA-GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAACTCGATTCCGATAGCCGGCTTTTCTCTATTTCSATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGA
ATATTGAAAAAACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTCCAACTATGGCAAGAAAAGTTTTTTCTCTT
TTTTTTCTAGAATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCAT
TTAGTTCAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTCAAAAAAAAAAATTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTT
ATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTTTACCAGGACTAACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCA
ATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCTACTAGAAAAACAAAAATTGCGGTTTCATTACGGACT
ORCR9                 ----AGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAA-------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
```

```
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
-----------------------------------------
OSRP1                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAACTCGATTCCGATAGCCGGCTTTTCTCTATTTCGATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGA
ATATTGAAAAAACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTT
TTTTTTCTAGAATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCAT
TTAGTTCAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTC-AAAAAAAAA-
TTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTTATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTTTACCAGGACT
AACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCAATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCTACTAGAAAAA
CAAAAATTGCGGTTTCATTACGGACT
OSRP5                 ----AGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
OSRP10                ----AGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
```

```
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGAACGAATCATTTACTCCACCGTCTGCTAT
ATCTTTTTAAGATCTGATGAATTGGATGAAAATAAAGATARAGTCCCATTCTACATGTCAATACCCACAACAATGAAATTTATTGTAAAAGGAA
AATCCKTCKACTTTAAAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAC------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
PRHP1                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATGGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAC------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
PRHP5                 ----AGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAA-------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
```

```
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
-----------------------------------------
PRHP8                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAACTCGATTCCGATAGCCGGCTTTTCTCTATTTCGATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGA
ATATTGAAAAAACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTT
TTTTTTCTAGAATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCAT
TTAGTTCAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTC-
AAAAAAAAAATTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTTATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTT
TACCAGGACTAACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCAATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCT
ACTAGAAAAACAAAAATTGCGGTTTCATTACGGACT
ROMT1                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAA-------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
-----------------------------------------------------------------
TCGATTCCGATAGCCGGCTTTTCTCTATTTCGATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGAATATTGAAAAA
ACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTTTTTTTTCTAGA
ATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCATTTAGTTCAATT
TTAGTTTAGGTTTATTCTCGAAAATGAAATTTC-
AAAAAAAAAATTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTTATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTT
TACCAGGACTAACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCAATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCT
ACTAGAAAAACAAAAATTGCGGTTTCATTACGGACT
ROMT3                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
```

```
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
ROMT5                   TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAAA-GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
SLMT1                   TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--RAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCAWARATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAARAARAAAGTTKGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAAKGGAATAATAACKGGAATAAAAAAGGAAACTATGGATTATTARACAAKGCATTTTTATGTTAT
GATTTCGGKGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAARAAAAAAAAAR-AA-C-
CCCTTATTTAAWTTGAAAKGAAYCCYCTTTTCCRAAYCYCATTAAMCYCCCMCTAAAAAA-GKGAACMCYCRRKTTTTC-AKGATT-CTTTT-
```

```
AGGAT-CCT-AT-C-TKGATTACS-CCCAACTCCCTKGTTTGACAAAARATCCWKGATT--AT-ATAAWA-ATTACATTGTASCGG-G--
TATAGTTTANG-GGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCATTGACCGAAT-
CGAGGAAAAACTCGATTCCGATAGCCGGCTTTTCTCTATTTCGATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGA
ATATTGAAAAAACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTT
TTTTTTCTAGAATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCAT
TTAGTTCAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTCNAAAAAAAAAATTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTT
ATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTTTACCAGGACTAACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCA
ATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCTACTAGAAAAACAAAAATTGCGGTTTCATTACGGACT
SLMT4                    TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------
SLMT8                    TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------
------------------------------------------------------------------------------
SURK1                    TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
```

```
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAA-------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
-----------------------------------------------------------------
TCGATTCCGATAGCCGGCTTTTCTCTATTTCGATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGAATATTGAAAAA
ACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTTTTTTTTCTAGA
ATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCATTTAGTTCAATT
TTAGTTTAGGTTTATTCTCGAAAATGAAATTTC-
AAAAAAAAAATTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTTATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTT
TACCAGGACTAACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCAATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCT
ACTAGAAAAACAAAAATTGCGGTTTCATTACGGACT
SURK3B                TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAC------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
SURK7                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
```

```
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
WHMT1                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--RAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCAWARATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAARAARAAAGTTKGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAAKGGAATAATAACKGGAATAAAAAAGGAAACTATGGATTATTARACAAKGCATTTTTATGTTAT
GATTTCGGKGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAARAAAAAAAAAR-AA-C-
CCCTTATTTAAATTGAAAKGAAYCCYCTTTTCCAAATCYCATTAAACYCCCMCWAAAAAA-GKGAACMCTCRRKTTTTC-AKGATT-CTTTT-
AGGAT-CCT-AT-C-TKGATTACS-CCCAACTCCCTKGTTTGACAAAARATCCWKGATT--AT-ATAAWA-ATTACATTGTASCGG-GG-
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTANCCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTANTTGATATTCCGATCAAAAACTTTAGTT-
CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAATTCGAGGAAAAACTCGATTCCGATAGCCGGCTTTTCTCTATTTCGATTCGATTTTT
TACATGATTCATGATAATATCCTTTTTTGAAATCAAACGAATATTGAAAAAACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTT
ATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTTTTTTTTCTAGAATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTAT
TTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCATTTAGTTCAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTCAAAAAAAAAA
ATTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTTATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTTTACCAGGAC
TAACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCAATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCTACTAGAAAA
ACAAAAATTGCGGTTTCATTACGGACT
WHMT4                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAAC-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
WHMT7                 ----AGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
```

```
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAC------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
WHSF1                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAACTCGATTCCGATAGCCGGCTTTTCTCTATTTCGATTCGATTTTTTACATGATTCATGATAATATCCTTTTTTGAAATCAAACGA
ATATTGAAAAAACGTTTCTCCTTTTCCAAAGATTACGGATTTTTTGATTATGTTATAGGAATGTTGAACTATGGCAAGAAAAGTTTTTTCTCTT
TTTTTTCTAGAATATAATAGAAAGAGTTTTGGATATTTCTAGAATTCATCTTATTTTTCTTTCTAACTTCCGTCGATTTGGCTTCATTTATCAT
TTAGTTCAATTTTAGTTTAGGTTTATTCTCGAAAATGAAATTTC-AAAAAAAAA-
TTAATATAAATTAAGCATAAGGAGTCTTTATGTCGCGTTATCGAGGACCTCGTTTCAAAAAAATACGCCGTCTGAGAGCAGCTTTACCAGGACT
AACTAAAAAATGGCCTAAACGCAGAGCTCGTAGAACCCAATCGCGTTTCGCGAAAAAGAAAAAATCTCAATATCATATTCGTCTACTAGAAAAA
CAAAAATTGCGGTTTCATTACGGACT
WHSF5                 TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
```

```
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCA-TGACCGAAT-
CGAGGAAAAC------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
WHSF10               TCAGAGAAACCCCGGAAAAAAAAAA-TGGGCAATCCTGA-
GCCAAATCCTGTTTTTTGCAAACAAACAAAGATTCCGAAAGC-
TAAAAAAAAAAGGATAGGTGCAGAGACTCAATGGAAGCTGTTCTAACAAACGGAGTTGAATGGATTGGTAGAGTAATAGAAATCCTTTTCTCGA
AACTTACGAAAAAATGAAA-GAAAACTCT--AT---T-C-TATA--TA--T---TATAGGTAATGAAATACTATA---T--CA-----
TTAATATAGAACAAAATATGTATCCCCATTTTTT-
ATAGCAAAAATGGAACAATTCTTGTGAATTGATTTCACATTGAATAAAAAATTGCATATTTATTGATCGAATCATTTACTCCACCGTCTGCTAG
ATCTTTTTAAGATCTGATGAATTGGATGAGAATAAAGATAGAGTCCCATTCTACATGTCAATACCGACAACAATGAAATTTATTGTAAGAGGAA
AATCCGTCGACTTTAGAAATCGTGAGGGTTCAATGACAAAAAAAGATAATTACTGTATTACATTACACGGAAAGAAAGGATTCCAAAAATTCTT
TCTTTCTCTCTTTATCT--AT-ATAT-TATA-TAG-ATA--TA---T-----A--
AGCACTTCCATTTCGATTAAGTTAGGACTGGAAAGGTCTAATAGAAACAAAAAAAAAA--GAATAATTCTTTGCTTTCATTTTGTACG-
AAAGGGTCTTTTTCTTTTTATTCCTCCGGTCGGGCTAGGTTAGTAACTAACCGGGCCCTTTTTTTGTTCCAACGAATCATAGATAAAACGATTT
ATCTAATTTGAAACCAAAAATACTTGCTATTAAAACAAGAAGAAAGTTTGATTTCTTATTATTCTTTCTT--TTTT--
ATTTACTTTTTTTAAGTATTTTAATACAAAATGGAATAATAACTGGAATAAAAAAGGAAACTATGGATTATTAGACAATGCATTTTTATGTTAT
GATTTCGGTGTTTTTATCCAACCGTATCTATCCAAATTCTTCCAGTAAAAGAAAAAAAAAG-AA-C-
CCCTTATTTAAATTGAAATGAATCCTCTTTTCCGAATCTCATTAAACTCCCACTAAAAAA-GTGAACACTCGAGTTTTC-ATGATT-CTTTT-
AGGAT-CCT-AT-C-TTGATTACG-CCCAACTCCCTTGTTTGACAAAAGATCCATGATT--AT-ATAATA-ATTACATTGTAGCGG-G--
TATAGTTTA-GTGGTAAAAGTGTGATT-CGTTC---TATTAACCCCCTTTAA--
TAGTTAAAGGGTCTTTCGGTTTAATTGATATTCCGATCAAAAACTTTAGTT-CTTAAAAGGATTAAATCCTTTACCTCGCAATGACCGAAT-
CGAGGAAAAC------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------
;
end;

begin sets;
  charset trnLF = 1-520;
  charset trnSG = 521-1398;
  charset rpS4 = 1399-1938;
  charsetpartition combined = trnLF: 1-520, trnSG: 521-1398, rpS4: 1399-1938;
end;
```