SAMPLES OF HOW TO INCLUDE PREVIOUSLY PUBLISHED ARTICLES
AS CHAPTERS IN A THESIS OR DISSERTATION

by

The Thesis Office

The Thesis Office
The University of Utah
2014
This packet contains samples of how previously published articles should be included in a dissertation or thesis. Please note that any chapter that has not been published when beginning the format approval process will need to be formatted according to the University of Utah requirements. For additional information on formatting requirements, please consult A Handbook for Theses and Dissertations found on the Thesis Office website.

The first sample in this packet is from a document created in Word. Note the following for Chapter 2, the previously published chapter:

- The Table of Contents lists all the first-level subheadings in the previously published chapter, just as it does for the other unpublished chapters.
- The title of Chapter 2 is exactly the same as the published article’s title.
- On the part-title page, there is a permission statement containing all pertinent information such as authors, article title, journal name, volume and page numbers where the article appeared.
- The page numbers for the dissertation appear in the top right corner; that is, the pages on which the previously published article has been inserted into the dissertation are still numbered consecutively with the rest of the dissertation pages.
- The article fits within the 1 ¼” side margins and 1” top and bottom margins.
# TABLE OF CONTENTS

ABSTRACT...........................................................................................................iii

ACKNOWLEDGMENTS.......................................................................................vii

Chapter

1. INTRODUCTION...............................................................................................1
   References.......................................................................................................5

2. PHYLOGENETIC RELATIONSHIPS IN SOLANUM SECTION ANDROCERAS
   (SOLANACEAE).............................................................................................7
   Abstract........................................................................................................8
   Introduction....................................................................................................8
   Materials and Methods................................................................................9
   Results........................................................................................................11
   Discussion.....................................................................................................13
   Acknowledgments.......................................................................................15
   Literature Cited...........................................................................................15

3. MOLECULAR DELIMITATION OF CLADES WITHIN THE NEW WORLD
   SPECIES OF THE “SPINY SOLANUM” (SOLANUM SUBGENUS
   LEPTOSTEMONUM)........................................................................................17
   Abstract.......................................................................................................18
   Introduction....................................................................................................18
   Materials and Methods................................................................................19
   Results..........................................................................................................20
   Discussion.....................................................................................................23
   Acknowledgments.......................................................................................28
   References.....................................................................................................28

4. A REVISION OF SOLANUM SECTION ERIOPHYLLUM.................................31
   Abstract.......................................................................................................31
   Introduction....................................................................................................31
CHAPTER 2

PHYLOGENETIC RELATIONSHIPS IN SOLANUM

SECTION ANDROCEAS (SOLANACEAE)

Phylogenetic Relationships in Solanum Section Androceras (Solanaceae)

Stephen R. Stern, Terri Weese, and Lynn A. Bohs

1Department of Biology, University of Utah, 257 South 1400 East, Salt Lake City, Utah 84112-0840, U.S.A.
2Author for correspondence (bohs@biology.utah.edu)

Communicating Editor: Anne F. Nowak

Abstract—The Leptospermum clade of Solanum contains approximately 350–450 species, including the cultivated eggplant, S. melongena. This clade is characterized by the presence of prickles and spicules at areole bases. Solanum section Androceras, the focus of this study, is a group of 12 species belonging in the Leptospermum clade. This section is unusual in the genus because of its mostly warm temperate distribution and diverse morphological, biochemical, and genetic characteristics. We undertook phylogenetic studies among 47 Solanum taxa, including 11 species and all varieties of sect. Androceras, using DNA sequence data from two nuclear regions (ITS and the 5S intron) and the mitochondrial gene trnF–psbA. The combination of morphological and molecular data supports sect. Androceras as a monophyletic group sister to Solanum sect. Chilopsis. Only one of the three genera present in previous taxonomic works (Solanum, Chilopsis, and Androceras) is supported by morphological data. A recent cladistic analysis of Solanum sect. Androceras (Ganguly et al. 2001) was updated to include additional species. Species-level relationships were also examined, and it was found that two species, S. leucostomum and S. citruloides, are not closely related. The ancestral flower color in sect. Androceras appears to be violet, with white and yellow flowers restricted to more derived clades. Character changes used to diagnose sect. Androceras, such as an inflorescence with many individual flowers, appear to have evolved more than once in the section.

Key words—phylogeny, morphometrics, ITS, DNA sequences.

Solanum L. (Solanaceae), thought to contain approximately 1,400 species, is one of the 10 largest genera of flowering plants (Frodin 2004, Bohs 2005). It also contains economically important species such as the tomato (S. lycopersicum L.), eggplant (S. melongena L.), potato (S. tuberosum L.). Recent studies of the genus range from sequencing the genome of the tomato (Mueller et al. 2005, http://www.sgn.cornell.edu/) to resolving phylogenetic relationships within Solanum as well as species level taxonomy (Knapp et al. 2004, http://www.nhm.ac.uk/solanumresources/). With respect to the phylogeny of the genus, analyses of DNA sequence data have helped to identify the major groups within Solanum, the largest of which is the Leptospermum clade (approximately 350–450 species) (Bohs 2005, Levin et al. 2006). This group is commonly known as the “spiny solanums” due to the presence of sharp spinae-like prickles.

Within the Leptospermum clade, Solanum sect. Androceras is unique in many features including distribution, flower and fruit morphology, and chemistry. Its morphological characteristics, speciﬁcally thorny morphology, are so distinct that Nattell (1818) described the species in the genus Androceras, although he noted the similarities between Androceras and Solanum. Marriott (1927) placed the species of Androceras into Solanum sect. Androceras. Whalen (1973a) provided a detailed revision of Solanum sect. Androceras, including 12 species and 10 varieties, and divided the section into three series (discussed below: Table 1) based on hair, flower, seed, and chemical characteristics as well as geographical distribution. Species in the sect. range from the southeastern U.S.A. to Mexico, with the highlands around Mexico City, the northern Chihuahuan Desert, and the west coast of Mexico as centers of diversity (Table 1). This section is one of the major groups within Solanum and has a primarily north temperate distribution. Within its range, species of sect. Androceras are weedy annual herbs or perennial from persistent woody roots. Many species grow in warm, semiarid to arid regions with unpredictable seasonal rainfall. Chromosome counts have been reported for all species in sect. Androceras, and there are diploids with 2n = 24 (Whalen 1973a).

Typical Solanum flowers are radially symmetrical with stamens coming from terminal pores. They are usually buzz pollinated, with pollen the targets of pollinators (e.g., bees). Species in sect. Androceras conform to this basic plan, but are further specialized in being bilaterally symmetrical. The stamens within a single flower are unequal in size, with four small, straight upper anthers and an elongate lower anther (heterostyly). This elongated, inward-curved lowermost stamen can be a different color than the other stamens and is opposed by a slender style of similar shape (Fig. 1a, b, c). The position of the style alternates between the right and left side of the flower along the inflorescence, resulting in "mirror-image" flowers (e.g., S. androceras, Fig. 1d). Flowers of sect. Androceras, specifically S. androceras, have been extensively observed in field and natural history studies with a focus on the unusual stamen dimorphic (Todd 1882, Harris and Kuch 1962, Bowers 1973, Jessen and Barrett 2002). The upper four small stamens provide the pollen that the bee uses for food, whereas the lowermost, elongated stamen acts as a pollinating stamen by depositing pollen on one side of the bee's abdomen where it cannot efficiently be removed (Bowers 1975, Vallejo-Marín et al. 2009). The alternating right- and left-handed flowers have been shown to have higher outcrossing rates than plants manipulated to have either straight styles or right-handed or left-handed flowers only (Jessen and Barrett 2002). This might be especially important in maintaining genetic diversity in sect. Androceras, where most tested species have been found to be self-compatible (Whalen 1973a).

Most species of Solanum have fleshy berries, whereas fruits in sect. Androceras are dry at maturity and tightly enveloped by a prickly, accrescent calyx (Fig. 1d). Whalen (1973a) showed that these species are a "claster" dispersal mechanism, also seen in other members of the Leptospermum clade, particularly those of dry habitats, in which the fruits remain on the plant and the calyx splits open, tearing the dry berry (Symon 1984, Knapp 2002). This then acts as a "claster," shaking loose the small seeds. The large number of seeds produced by a single plant, in some cases over 5,000 seeds from an individual, corresponds to the observation that Solanum sect. Androceras is typically a weedy, colonizing group of species.

Some species of sect. Androceras have a unique suite of flavonoid compounds, such as 8-hydroxyflavonoids and C-glycosylflavonoids, not found in other Solanum groups (Whalen 1973a). Differences also exist in the chemical profiles between the three series within the section recognized.
Table 1. Species of Antidesma sect. Androcnus, including the series and their distribution according to Whalen (1979a). All taxa except S. leucocarpum were sampled in this study.

<table>
<thead>
<tr>
<th>Series Antrocnus</th>
<th>Species</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. angustifolium</td>
<td>Mill</td>
<td>Tropical Mexico south to Honduras</td>
</tr>
<tr>
<td>5. subg. Molinae</td>
<td>Mill</td>
<td>Distrito Federal, Hidalgo, and Mexico States with collections from Cuernavaca and the Sierra Madre, Mexico.</td>
</tr>
<tr>
<td>5. malaguetae</td>
<td>Dunal</td>
<td>Endemic to eastern Durango State, Mexico.</td>
</tr>
<tr>
<td>5. pseudovagum</td>
<td>Whalen (1979a).</td>
<td></td>
</tr>
<tr>
<td>5. aurantiacum</td>
<td>Whalen</td>
<td>Known only from the type locality in western Puebla, Mexico.</td>
</tr>
<tr>
<td>Series Fruticosum Whalen</td>
<td>5. ciliatifolium A. Braun var. ciliatifolium</td>
<td>Whalen (1979a).</td>
</tr>
<tr>
<td>5. var. ciliatifolium</td>
<td>Whalen (1979a).</td>
<td></td>
</tr>
<tr>
<td>5. var. acuminatum</td>
<td>Whalen (1979a).</td>
<td></td>
</tr>
<tr>
<td>5. var. ciliatifolium</td>
<td>Whalen (1979a).</td>
<td></td>
</tr>
<tr>
<td>5. var. aurantiacum</td>
<td>Whalen (1979a).</td>
<td></td>
</tr>
<tr>
<td>5. var. ciliatifolium</td>
<td>Whalen (1979a).</td>
<td></td>
</tr>
<tr>
<td>5. var. acuminatum</td>
<td>Whalen (1979a).</td>
<td></td>
</tr>
<tr>
<td>5. var. aurantiacum</td>
<td>Whalen (1979a).</td>
<td></td>
</tr>
</tbody>
</table>

by Whalen (1979a), such as the presence of methylolated aglycones, 8-hydroxyflavonoids and various flavones in sect. Veracerrum and Pacifico are flavones with clavatoxe:type B rings in sect. Pacifico and the presence of 8-oxogenated flavonoids in sect. Veracerrum (Whalen 1979a).

Although Whalen (1979a) revised sect. Androcnus and included a cladistic analysis based on 14 morphological and chemical traits, to date there have been limited molecular phylogenetic studies of this section. Two species of sect. Androcnus, S. molinae and S. ciliatifolium, were included in molecular phylogenies of the entire Leptospermum clade and were strongly supported as sister taxa (Levin et al. 2006). These studies place sect. Androcnus sister to sect. Cudrania with moderate support (84% bootstrap and 1.0 posterior probability in Levin et al. 2006). This relationship had not previously been proposed due to the fact that sect. Cudrania is a South American group of large shrubs and trees with fruits that may reach 10 cm in diameter and large flowers that are not heterostylous. A close relationship between sect. Androcnus and S. molinae is consistent with the monophyly of sect. Cudrania as proposed in the past due to their similar leaves, inflorescences, and accrescent calyces (Dunal 1813, 1852; Walpers 1841, Daneri 1973, Whalen 1979a, Lister et al. 2004). Both Whalen and Bohl (2000) found that S. molinae is sister to a clade composed of sect. Androcnus and sect. Cudrania. Whalen (1979a) noted that sect. Tetramerum (Venten.) Wapfl. as the sister group to sect. Androcnus, based on morphological similarities, but molecular studies unequivocally place the members of sect. Tetramerum quite distant from sect. Androcnus (Levin et al. 2006, Bohl et al. 2007, Whalen 1979a). While these studies provide hypotheses about relationships between sect. Androcnus and other Solanum sections, they did not extensively sample from within the section.

In this paper we use molecular phylogenetic methods to 1) test the monophyly of sect. Androcnus as currently circumscribed, 2) examine the phylogenetic relationships of sect. Androcnus with closely related members of the Leptospermum clade, 3) test the monophyly of Whalen's (1979a) series and species within sect. Androcnus, and 4) examine selected species-level relationships to test hypotheses of character evolution and speciation proposed by Whalen (1979a).

Materials and Methods

Tissue Sampling.—Eleven of the 12 species and all 10 varieties in sect. Androcnus were sampled for this study (Table 1). We were unable to obtain high-quality genetic DNA for the remaining species, which is known only from the type locality in Puebla, Mexico, due to a lack of available herbarium material. Specimens were determined using keys found in Whalen (1979a). Almost all of the specimens determined by the key Michael D. Whalen himself (indicated with asterisks in Appendix 1). We also included six members of sect. Cudrania as well as S. molinae, both shown by previous molecular studies to be closely related to sect. Androcnus (Levin et al. 2006, Bohl et al. 2007). Five other more distantly related species from the Acaciaphyllum and Batracophyllum clades of the Leptospermum clade were included to ensure sufficient outgroup sampling, and the tree was rooted using S. molinae, an even more distantly related Solanum from outside the Leptospermum clade. The final data set included 43 accessions, representing 11 natural species of sect. Androcnus as well as 12 outgroup species. All taxa, along with voucher information and GenBank accession numbers, are listed in Appendix 1.

DNA Extraction, Amplification, and Sequencing.—Total genomic DNA was extracted from fresh, air-dried, or herbarium material using the Qiagen plant mini-extraction kit (Qiagen Inc., Valencia, California). Amplification for each gene region followed standard procedures described in Taberlet et al. (1991), Bohl and Olmstead (2001), and Bohl (2004) for the trnL-F and trnL-F intergenic spacer regions. Levin et al. (2004) for rps16 and Lavin et al. (1998) for ITS. The ITS region was amplified as a single fragment using primers ITS1 and ITS4 (White et al. 1990) using PCR conditions described in Bohl and Olmstead (2001). When possible, trnL-F and rps16 were amplified as single fragments using primers a and b of the trnL-F (Taberlet et al. 1994) and primers scayF4 and scayF5 for rps16 (Lavin et al. 2005). Amplification conditions for the trnL-F followed Bohl and Olmstead (2001), conditions for rps16 followed Lavin et al. (2005). When necessary, overlapping fragments were amplified and assembled, using primers a with d, and c with b to amplify trnL-F, and primers scayF4 with E73R and TRN98 with 2R to amplify rps16. Specimens not amplifying for rps16 were amplified in
even smaller fragments using primers wxyF and the newly developed
EMR (5'-CAACATTTCAACCTAAG)-3') for the first fragment, the new-
primer EMW (5'-GTCATGCTCGGACATCTCT-3') and 1171R for the
second fragment, primers 1100F and 3'N (Perzl and Spanner 2001) for
the third fragment, and primers SF (Miller et al. 1999) and 28 for the final
fragment.
Amplification products were cloned using the Promega Wizard SV
PCR Clean-Up System (Promega Corporation, Madison, Wisconsin). The
University of Utah DNA Sequencing Core Facility performed sequencing
on an ABI automated sequencer. Sequences were edited in Sequencher
(Gene Codes Corp., Ann Arbor, Michigan) and all new sequences were
submitted to GenBank.

Morphological Data—The data matrix presented in Whalen (1979a,
Table 6), representing 11 morphological, 12 chemical, and one enzyme
character for species in sect. Androceras was added to the combined molecu-
lar data matrix with characters for outgroup species coded as missing data.

Sequence Alignment and Analysis—Sequence alignment for all
gene regions was straightforward and performed visually using Se-Al
(Rambaut 1996). The aligned datasets and representative phylogenetic
trees are available in TreeBASE (study number SB642).

Phylogenetic Analyses—Maximum parsimony (MP) analyses were
performed on each dataset separately and on the combined dataset both
with and without morphological data using PAUP*4.0 (Swofford 2002). All
characters were weighted equally in analyses that implemented
the function reconstruction (TRI) branch swapping with 1,000 heuristic
random addition replicates, each limited to 1,000,000 steps per repli-
cate. Gaps were treated as missing data. Bootstrapging (BS; Felsenstein 1985)
was used to evaluate branch support with 1,000 random addition
replicates and the TRI branch swapping limited to 1,000,000 steps per repli-
cate. Datasets were further analyzed using TNT (Goloboff et al. 2008)
to search for shorter trees than were obtained in standard PAUP analy-
ses. Congruence of the datasets was tested using partition homogeneity
tests (incorrigible length difference test [ILD]), Farris et al. 1994, 1996).
implemented in PAUP*. One thousand heuristic partition homogeneity
replicates were completed, each with 10 random-addition sequence replicates,
TBR branch swapping, M ulttree off, and gaps treated as missing data.

Bayesian Analyses—Prior to Bayesian analysis (BI), a general model of nucleotide evolution was selected for each of the separate and com-
bined datasets using the ACE criterion identified in ModelTest 3.7 (Posada
and Crandall 1998). MrBayes 3.1 (Ronquist and Huelsenbeck 2001) was
used to analyze the individual and combined datasets. For each analysis
20 replicates were run for four Markov chains, each initiated from a ran-
donut tree and sampled every 1,000 generations using the step rule to stop
the analysis when standard deviations between the trees reached 0.01. All
parameters from each analysis were visualized graphically and the sam-
pled obtained prior to achieving stationarity were discarded as a burn-in.

Constrained Analyses—Constrained trees were constructed in MacClade
4.0 (Maddison and Maddison 2000) to constrain each of Whalen’s (1997b)
series as monophyletic, 2) only the taxa in set Androcreea as monophyletic,
3) only the taxa in set Violaceasporum as monophyletic, and 4) yellow-
deciduous taxa as monophyletic. Parsimony analyses were performed with
the constraint enforced using TBR branch swapping with 1,000 heuristic
random addition replicates, each sampled from 1,000,000 steps per rep
licate. Trees were then compared with the most parsimonious trees using

RESULTS

Phylogenetic Analyses—Descriptive statistics for the molecular
datasets and phylogenetic analyses for the 43 accessions are
given in Table 2. Missing data comprised 0.00897% of the
combined data matrix (148 bases from a total of 171,707). For
the individual datasets, the ITS-T-F region yielded the least
resolved phylogeny in both MP and BI analyses. The most
data produced the most resolved trees with the highest num-
ber of strongly supported ingroup nodes (Table 2). In general,
the parsimony strict consensus and BI majority rule consensus
trees from the combined dataset differed only in the degree of
resolution, with BI topologies more resolved than parsimony
trees (Table 2). Clades with low posterior probabilities
(PP) in BI analyses were often collapsed in MP strict consen-
sus trees (individual trees not shown).

More nodes were strongly supported by combining the
three datasets than were obtained in any of the separate
analyses (Table 2, Fig. 2). Inclusion of morphological data did not
affect either the topology or resolution of the phylogeny com-
pared to the combined molecular dataset analyzed alone. The
only differences between these and the strictly molecular
trees were slight differences in support values for a few nodes.

Topological Conflicts—According to the results of the ILD
tests, the three data partitions in the combined data set were
found to be incongruent (p = 0.033), so pairwise ILD tests were
run. The nuclear datasets (ITS and 18S) were found to be
more congruent (p = 0.21) as were the rDNA and T-F datasets
(p = 0.071). The incongruence of the datasets is likely due to
the disparity in the size and substitution rates of the different
datasets (Dolgin et al. 2000, Barker and Lovette 2002; Darlu
and Leconte 2002). However, with few exceptions, each
DNA sequence region consistently identified the same major
well-supported clades comprising identical species groups,
but relationships among clades were often not strongly sup-
ported (BS values < 90%), or were unresolved, and thus
cannot be considered conflicting under Wiens’ (1998) criteria.
The BI analyses gave more conflicting nodes (cut-off < 0.05 PP),
but posterior probabilities are known to be inflated relative to
bootstrap values (Cummings et al. 2003; Ericksen et al. 2003;
Simmons et al. 2004). Our discussion will be focused on the
topology of the BI majority rule and MP strict consensus trees
based on combined molecular data (Fig. 2).

Phylogenetic Relationships—Sectional Relationships and
Monophyly of Section Androcreea—all data sets
strongly support the monophyly of sect. Androcreea as circumscribed
by Whalen (1997b, 1997c, 1998). 100% BS, 1.0 PP in ITS, 100% and
gene trees and 99% BS, 1.0 PP in ITS-T-F.

Although not supported in the single-gene analyses, the combined
dataset supports sect. Cunia as sister to sect. Androcreea (85% BS, 1.0 PP),
with Violaceasporum sister to the clade composed of sects. Androcreea and
Cunia (85% BS, 1.0 PP).

Monophyly of the Series Within Section Androcreea—of
the three series identified by Whalen (1997b), the phylogeny
supports only sect. Pacificum as a monophyletic group, termed
the Pacificum clade in Fig. 2. This relationship is supported in
the individual ITS (87% BS, 1.0 PP) and 18S datasets (68% BS,
1.0 PP) but not in the ITS-T-F dataset: the combined dataset
resolves this group with 100% BS and 1.0 PP. Three of the five
species of sect. Androcreea form a moderately strongly sup-
ported Rosaceae clade composed of S. rosaceum, S. frutico-
to com and S. argentifolium in the only (82% BS, 1.0 PP)
and combined trees (94% BS, 1.0 PP). Violaceasporum (sect.
Androcreea) is not placed in the ITS, ITS-T-F, and combined
analyses; the only analysis places this species as sister to the
Pacificum clade with moderate support (86% BS, 1.0 PP).
The total number of Whalen’s sect. Androcreea, S. holbro-
sonianum, is moderately supported (82% BS, 1.0 PP) as sister
to a large clade of species, placed by Whalen (1997b) in sect.
Violaceasporum, in the combined analyses; but the relationship
is not recovered in any of the individual analyses. Whalen’s
sect. Violaceasporum is clearly polyphyletic, with a large clade
composed of S. heterodoxum var. setigermis, S. atrinervium var.
ciliarifolium and setigermis, and S. denticulatum forming a
monophyletic group, here termed the Setigermis clade, in the
only (80% BS, 1.0 PP) and combined analyses (92% BS, 1.0 PP;
Fig. 3). The remainder of the taxa belonging to Whalen’s
sect. Violaceasporum, including S. fruticosum, S. ciliarifolium var.
knobchati, and S. heterodoxum var. heterodoxum and renovero-
atum form a grade at the base of the Androcreea clade in the
combined analyses: "Elder 40", a potentially undescribed

Table 2. Descriptive statistics for the datasets analyzed. Strongly supported nodes for parsimony indicates those with a 100% BS. Bayesian strongly
supported nodes are those with 0.95 PP.

<table>
<thead>
<tr>
<th>Data Partition</th>
<th>Average Sequence Length</th>
<th>Number of Ingroup Taxa</th>
<th>Number of MP Bas</th>
<th>Tax Length</th>
<th>CI</th>
<th>Number of Strongly Supported Nodes</th>
<th>Parsimony Ingroup nodes</th>
<th>Model Selected</th>
<th>Number of Strongly Supported Nodes</th>
<th>Parsimony Ingroup nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITS</td>
<td>584</td>
<td>121</td>
<td>13,691</td>
<td>431</td>
<td>0.688</td>
<td>0.789</td>
<td>11 (6)</td>
<td>GTR + I + G</td>
<td>21 (15)</td>
<td></td>
</tr>
<tr>
<td>18S</td>
<td>1,731</td>
<td>165</td>
<td>48</td>
<td>428</td>
<td>0.844</td>
<td>0.895</td>
<td>17 (12)</td>
<td>GTR + I + G</td>
<td>32 (26)</td>
<td></td>
</tr>
<tr>
<td>rDNA</td>
<td>2,048</td>
<td>68</td>
<td>52,750</td>
<td>108</td>
<td>0.916</td>
<td>0.895</td>
<td>6 (3)</td>
<td>GTR + I + G</td>
<td>13 (14)</td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>4,285</td>
<td>340</td>
<td>10</td>
<td>1,085</td>
<td>0.733</td>
<td>0.829</td>
<td>18 (12)</td>
<td>GTR + I + G</td>
<td>36 (26)</td>
<td></td>
</tr>
<tr>
<td>Combined + Morphological</td>
<td>4,600</td>
<td>354</td>
<td>6</td>
<td>1,127</td>
<td>0.728</td>
<td>0.828</td>
<td>16 (12)</td>
<td>GTR + I + G</td>
<td>38 (26)</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 2. 50% majority rule tree from the Bayesian analysis of the combined dataset. Numbers above branches are bootstrap values over 50%. Numbers below branches are posterior probabilities from Bayesian analysis. Branches that collapse in the parsimony strict consensus tree but are present in the Bayesian majority rule tree are shown as dashed lines. Species of sect. Androceras placed by Whalen (1974b) in sect. Setigerae are in bold, species in sect. Pacificum are underlined, and in sect. Villosum/Winnem are in non-bold italics. Solanum "Elder 46" was not placed in any of Whalen’s (1974b) series, see text for discussion. The clades discussed in the text are labeled.
Species-level monophyly was examined in a number of taxa with multiple accessions sequenced in the phylogeny. In the cases of S. aequilabium, S. youngianum, S. psilophyllum, S. tribuloides, and S. youngianum, all accessions of the same species formed monophyletic groups with strong support in the combined trees. Furthermore, the multiple accessions sequenced of S. aequilabium, S. youngianum, and S. tribuloides each emerged as monophyletic in all combined analyses, but S. youngianum, S. youngianum, and S. youngianum, were not supported as monophyletic; as multiple accessions of these taxa did not group together in the combined analyses.

Constraint Analyses—Constraint of all of Whalen's series to be monophyletic resulted in trees significantly different than the most parsimonious tree from the combined dataset (Templeton's test, p = 0.001). When constraining sect. Andrucea and Volucelopodium individually, the trees were also significantly different than the most parsimonious tree from the combined dataset (p = 0.0055 and 0.0477, respectively).

Trees containing all of the yellow-flowered taxa (i.e., species of sect. Andrucea minus S. tribuloides) were not significantly different than nonconstrained trees (Templeton's test, p = 0.0009).

Discussion

Sectional Relationships and Monophyly of Section Andrucea—Despite the various hypotheses regarding the sister group to sect. Andrucea, our data support previous molecular studies in finding sect. Creminum as sister to sect. Andrucea (Levin et al., 2003). These groups are morphologically distinctive and thus relationship merits further study. Solanum asyntrichum is sister to a clade composed of sect. Andrucea and sect. Creminum despite the fact that S. asyntrichum and sect. Andrucea share highly divided leaves and strongly accrescent calyces, characters not found in sect. Creminum. Lester et al. (2000) also found the seeds of sect. Andrucea and Cryptopogon, to which S. asyntrichum belongs, to be remarkably similar. We were not able to sample other members of sect. Cryptopogon but further sampling might possibly place this group sister to sect. Andrucea.

All three data sets strongly support the monophyly of sect. Andrucea as circumscribed by Whalen (1979a, 1984). This molecular evidence, combined with unique morphological traits found in the leaves, flowers, and fruits, its distinctive flavonoid chemistry, and geographical distribution, leave little doubt that Solanum sect. Andrucea is a monophyletic group.

Character Evolution and Morphology of Whalen’s Series in sect. Andrucea—Whalen’s (1979a) three series within sect. Andrucea were distinguished by utricle, flower, and seed morphology as well as flavonoid chemistry and geographical distribution (Table 1). In Whalen (1979a), circumscribed these series as natural phylectic groups; however, they were not defined in strict monophyletic terms (see paraphyly of sect. Andrucea and Volucelopodium in Fig. 15 in Whalen 1979a).

It is clear in examining his matrix of morphological characters (Table 6 and Fig. 15 in Whalen 1979a), that many are homoplasious or autapomorphic. Additionally, the assessment of ancestral and derived characters as well as coding of characters are based on the author’s interpretations (see secondary lost characters in Table 5 in Whalen 1979a) and could be differently interpreted by other taxonomists. Given this and the fact that our combined molecular dataset contains 240 parsimony informative characters, it is not surprising that the addition of the 14 characters from Whalen (1979a) dataset does not change the topology or resolution of the phylogeny (results not shown). The few synapomorphic characters in Whalen’s (1979a) character matrix show support for sect. Pacificum, the only one of the three series that emerges as a monophyletic group in our molecular trees. Characters unique to this series include white, deep yellow corolla, slightly protruding styles, and a geographical center of distribution on the Pacific slope of the Sierra Madre Occidental on the west coast of Mexico. Apparently these characters arose once in the Pacificum clade, although confirmation of this awaits sampling of the third member of sect. Pacificum, S. youngianum.

Neither sect. Andrucea nor Volucelopodium is supported as monophyletic in the molecular analyses. These series were paraphyletic in Whalen’s (1979a) cladistic analysis and the nonmolecular characters that supported these groups are largely convergent. For instance, sect. Andrucea was characterized by Whalen (1979a) as having yellow or multilugulate cuneate hays and yellow corollas, lacking flavonoid compounds found in the other two series, and a distribution centered in the central Mexican highlands around Mexico City. These characters are found in species of the Rostratum clade (Fig. 2), but also in S. youngianum, which does not form a part of this clade. Conversely, Whalen (1979a) placed S. tribuloides into sect. Andrucea despite its pale blue or white corollas. Support and resolution along the backbone of the tree obtained here is weak or lacking, precluding firm conclusions about character evolution in sect. Andrucea based on the more all-parsimonious trees. However, constraining all three series each to be monophyletic as well as constraining the taxa of Whalen’s sect. Andrucea and Volucelopodium individually to be monophyletic resulted in trees significantly different than the most parsimonious trees from the combined dataset. This further indicates that these two series are likely nonmonophyletic and that the characters that Whalen proposed to diagnose them have evolved multiple times. On the other hand, when all yellow-flowered taxa (i.e., species of sect. Andrucea minus S. tribuloides) were constrained to monophyly, the constrained trees were not significantly different than nonconstrained trees. Therefore, the hypothesis of a single origin of yellow corollas within sect. Andrucea cannot be rejected.

According to Whalen (1979a), nine of the species of sect. Andrucea are recognized annual herbs with wide edaphic tolerances Solanum microbotrys, S. youngianum, and S. tribuloides, however, are calciphilic herbaceous perennials; judging from their widely separated positions on the molecular trees, it appears that the latter traits evolved independently in the three species.

Biogeographical Relationships—Based on his interpretations of cladistic relationships in sect. Andrucea, Whalen
(1970a) considered sect. Androcuis to be pleomorphic within the section, implying an origin for the section in the central Mexican highlands (Whalen 1970a, 1983). However, the molecular phylogenies place S. truxpna, included in sect. Androcus, as the sister to the remainder of sect. Androcuis with good support (82% BS, 1.0 PP). Solanum truxpna occurs in the northern Chihuahua Desert near the Texas-Mexico border, pointing to a more northern origin for the section. In the BI trees, the Glendalosa clade is in turn sister to the remainder of the species (Fig. 2). Species of this clade are also found in the northern Chihuahua Desert and range into the southwestern U.S.A., consistent with a northern origin. However, this latter relationship is poorly supported and collapses in the MP strict consensus trees. Nonetheless, molecular evidence relates Whalen's (1970a, 1983) hypothesis of a central to southern Mexican origin for sect. Androcuis.

Clades Within sect. Androcuis—Rostatrum clade—The Rostatrum clade contains S. rostratum, S. fructose-toto, and S. angustifolium, three of the five species placed by Whalen (1970a) in sect. Androcuis. Solanum rostratum is a widely introduced weed and common in the central and western U.S.A., but Whalen (1970a) considered central Mexico to be its area of origin due to the high level of morphological variability in this region and because many of the sister taxa proposed by Whalen (1970a) occur there. Our phylogeny samples accessions from both the U.S.A. and Mexico and also all form a strongly supported group. Combined with many morphological characters, there is little doubt that, although it is the most widespread species in the section, S. rostratum is a monophyletic and distinct species. The other members of the Rostatrum clade have more restricted distributions. S. fructose-toto is found in the vicinity of Mexico City and Ciudad Durango, and S. angustifolium is found from southern Mexico through Honduras. Although S. fructose-toto is vegetatively similar to S. rostratum, Whalen did not encounter hybrids or collections intermediate between the two species in reproductive characteristics. Therefore, he states that the overlap in vegetative characteristics between the species probably represents natural variation. Whalen (1970a) considered S. angustifolium to be closely related to S. rostratum but also called it a bridging taxon between the two. Androcuis and Violafoayan. Our phylogeny indicates that, despite sharing trichome and flavonoid characters with species in Whalen's sect. Violafoayan, S. angustifolium is an early clay closely related to S. rostratum.

Pitocinum clade—The Pitocinum clade is found in western Mexico along the Pacific slope of the Sierra Madre Occidental and inland in central Mexico. This clade comprises two of the three species placed by Whalen (1970a) in sect. Pitocinum, S. brevifolium, the third, was not sampled. Solanum grayi has been divided into two varieties based on flower size. The small-flowered form is known as S. grayi var. grayi, whereas the large-flowered plants are segregated as S. grayi var. grandiflorum. Our phylogeny sampled species from throughout the range of S. grayi and did not consistently separate these varieties. These varieties seem to have arisen from character displacement in areas where S. grayi occurs sympatrically with its purported sister species S. humboldtianum. Whalen (1970b) showed that S. humboldtianum and S. grayi have similar sized flowers over their distinctive ranges, but show strong character displacement where their ranges overlap in Sonora and northern Sinaloa, with the flowers of S. grayi much smaller than in other parts of its range. Solanum grayi and S. humboldtianum were shown to successfully hybridize in experimental crosses, but Whalen (1978, 1979) posits mechanical isolation via character displacement of floral traits in areas where the two species overlap, indicating that in nature they would not share the same pollinators and would effectively be reproductively isolated. Although our phylogenetic data suggest that the varietal distinctions in S. grayi might not be warranted, additional sampling from this species is needed to examine this question. The final member of Whalen's sect. Pitocinum, S. loweianum, is a rarely collected species and thus material was not available for this study. It is endemic to western Puebla and is morphologically similar to S. grayi, thus would likely be included in the Pitocinum clade.

Setigeroid clade—The Setigeroid clade is strongly supported in our phylogenies and contains S. davisense, S. crassifolium var. crassifolium and setigerum, and S. heterodoxum var. setigeroides. These species all occur in the southwestern U.S.A. and the area along the Texas-Mexico border. Our phylogeny shows that S. davisense is closely related to S. crassifolium var. crassifolium (76% BS, 1.0 PP), a result supported by allozyme data from Whalen (1970b). Solanum davisense is distinct from the other species of the Setigeroid clade due to its erect habit, acutely lobed leaves, smaller flowers, and smooth unligned seeds as well as chemical differences (Whalen 1970a). Divergence of S. davisense and S. crassifolium was likely due to the slight geographical separation of S. davisense at the margin of the range of S. crassifolium var. crassifolium (Whalen 1970a, 1979a). Solanum crassifolium var. crassifolium and setigerum do not form a monophyletic group in either the MP or BI combined analysis. Morphology of S. crassifolium var. crassifolium and S. setigerum is deviant from the monophyletic nature model, yet it receives strong support (1.00 PP) in the BI 50% majority rule tree. Therefore, it is unclear whether the two varieties should be recognized as taxonomically distinct entities. As indicated by the common varietal name setigerus (Latin for "bristly"), S. crassifolium var. setigerum and S. heterodoxum var. setigeroides share morphological similarities and have also been found to have a history of hybridization (Whalen 1970a). This, combined with the phylogenetic relatedness of these taxa, warrants a more detailed taxonomic investigation of these species and varieties to determine the relationship and specific delimitations of members of the Setigeroid clade.

Glandulosa clade—The Glandulosa clade presents interesting taxonomic and biogeographic problems. This clade contains S. crassifolium var. knoblichii, S. heterodoxum var. novomexicanum and an undescribed species here called "Elder 48" based on the collector and collection number. Solanum crassifolium var. knoblichii morphologically resembles var. crassifolium but is restricted to western Chihuahua state in Mexico. It has longer hairs and more spreading fruit pedicels than the other varieties of S. crassifolium but, due to a lack of collections, other morphological differences are not apparent. It is distantly related to its conspecifics, which occur in the Setigeroid clade (see above), and deserves further collection and taxonomic study. Solanum heterodoxum var. novomexicanum was given specific status as S. novomexicanum (Barrett) Wooten & Standley (1913). Despite the large geographic separation between S. heterodoxum var. heterodoxum from the area around Mexico City and var. novomexicanum from New Mexico, Whalen (1970a) felt that these varieties resembled each other except for the more stellate corollas in the latter variety. The geographically close S. heterodoxum var. setigeroides occurs in adjacent areas of New Mexico, Arizona, and the Texas-Mexico border. This variety
is distinct morphologically, with densely prickly stems and much finer spines than the other varieties of S. heterodoxum. Given the distinct morphological traits and the phylogenetic distance between S. heterodoxum var. nevacuum and the other varieties, the specific classification of Wootten and Standley (1913) should be reconsidered. The final member of this clade, "Elder 46," is a collection from Jeff Davis County, Texas. This specimen has previously been identified as S. junipero var. grandifolium, S. advenus, and S. heterodoxum, but does not fit any of these species concepts. Whalen did not examine this specimen, and use of his key and comparison to specimens he annotated does not result in a satisfactory determination. Since it appears that many of the species in the section are restricted endemics, it is possible that this collection represents an undescribed species.

The three species in the Gladiolus clade share some morphological characteristics including a diminutive weakly annual habit, violet or occasionally white flowers, and simple, often glabrous leaves. Whalen (1979a) notes that the florescence profile for S. heterodoxum var. nevacuum is identical to that of var. nevacuum; however, the other members of the Gladiolus clade have not been sampled. The species in the Gladiolus clade have geographic ranges that do not appear to overlap, with "Elder 46" occurring in Jeff Davis County, Texas, S. heterodoxum var. nevacuum occurring in north central New Mexico, and S. grandifolium var. nevacuum restricted to Chihuahua, Mexico. Further systematic study and field collections will help to clarify the number of distinct taxa represented within this clade.

Solanum juniperum—The two accessions of S. juniperum emerge as a monophyletic group. This species has a very restricted range in the Durango state of north-central Mexico and has often been identified as S. juniperum. However, Whalen (1979a) cites many morphological differences as well as reproductive isolation as evidence that S. juniperum and S. juniperum are distinct species. Our phylogeny supports this separation, but there is little support for the relationship of S. juniperum with any of the other clades within sect. Atriplexes.}

Solanum rhomboideum—Solanum rhomboideum shares purple flower color with members of the Whalen's sect. Vidalichunus, but he placed it in a sect. Atriplexes due to geographical distribution, chemical characteristics (notably a lack of 8-hydroxyflavonoids and various flavonoids that are found in sect. Vidalichunus) and morphological features such as stellate trichomes and smooth seeds. Results from the combined analyses indicate that S. rhomboideum is more closely related to the other purple-flowered taxa that have been placed in the Seriptoglochum clade than to the species of Whalen's sect. Atriplexes, which include S. retusum, S. pruinosum, and S. angustifolium (Roststrum clade) as well as S. abominans.

Solanum heterodoxum var. heterodoxum—The position of S. heterodoxum var. heterodoxum within the section is unresolved and it is not placed with either of the other S. heterodoxum varieties. This isolated phylogenetic position mirrors its geographical distribution; Solanum heterodoxum var. heterodoxum occurs in the southwestern U.S., and northern Mexico, whereas var. heterodoxum is greatly disjunct in central Mexico. Solanum heterodoxum var. heterodoxum has less prickly stems with much shorter prickles than those of var. retusum and flowers with much more interpetiolar tissue than those of var. nevacuum. These differences, along with the phylogenetic results, indicate that S. heterodoxum as currently defined is almost certainly not monophyletic.
The second and final sample in this packet is from a document created in LaTeX.

Note the following for Chapter 3, the previously published chapter:

- The Table of Contents lists all the first- and second-level subheadings in the previously published chapter, just as it does for the other unpublished chapters.

- In the Table of Contents, the subheadings of the previously published chapter are numbered with a local numbering scheme even though the subheadings are not numbered in the actual article. A local numbering scheme is one in which subheadings, tables, and/or figures are numbered with the chapter number and then consecutively within that chapter; for example, 3.2 would indicate a subheading, table, or figure from Chapter 3 and that the subhead, table, or figure was the second one in Chapter 3, thus 3.2.

- In the List of Tables, those tables in the previously published article are also given a local numbering scheme, just like the subheadings. When a List of Tables or List of Figures is used and/or required and there is a previously published article in the thesis or dissertation, a local numbering scheme for tables and figures must be used.

- The entries in the List of Tables match the first full sentence of the table titles in the previously published chapter word for word.

- The title of Chapter 3 is exactly the same as the published article’s title.
• On the part-title page, there is a permission statement containing all pertinent information such as authors, article title, journal name, volume and page numbers where the article appeared.

• The page numbers for the dissertation appear in the top right corner; that is, the pages on which the previously published article has been inserted into the dissertation are still numbered consecutively with the rest of the dissertation pages.

• The article fits within the 1 ¼” side margins and 1” top and bottom margins.
# CONTENTS

**ABSTRACT** ................................................................. iii

**LIST OF TABLES** ......................................................... viii

**ACKNOWLEDGMENTS** .................................................... x

**CHAPTERS**

1. **INTRODUCTION** .................................................. 1
   1.1 Motivation .................................................... 1
   1.2 Longitudinal Studies .......................................... 4
      1.2.1 Objective of Longitudinal Studies .................. 4
      1.2.2 Properties of Longitudinal Studies ............... 6
      1.2.3 Clinical Applications ............................... 6
      1.2.4 Traditional Approaches ............................ 9
   1.3 Contributions ................................................ 10
   1.4 Overview of Chapters ..................................... 11

2. **FRAMEWORK FOR MODELING GROWTH TRAJECTORIES OF EARLY BRAIN DEVELOPMENT** ............... 12
   2.1 Introduction ................................................ 12
   2.2 Growth Models .............................................. 12
      2.2.1 Exponential and Monomolecular ............... 13
      2.2.2 Logistic ............................................. 14
      2.2.3 Gompertz ............................................ 14
      2.2.4 Selection of Growth Models .................... 14
   2.3 Mixed Effects Model ...................................... 15
      2.3.1 Linear Mixed Effects Model .................... 17
      2.3.2 Nonlinear Mixed Effects Model ............... 18
   2.4 Experiments and Results .................................. 19
      2.4.1 Model Selection .................................... 19
      2.4.2 Evaluation of Growth Curves .................... 21
      2.4.3 Hypothesis Testing ................................ 27

3. **REGIONAL CHARACTERIZATION OF LONGITUDINAL DT-MRI TO STUDY WHITE MATTER MATURATION OF THE EARLY DEVELOPING BRAIN** ............... 29
   3.1 Introduction ................................................. 30
   3.2 Materials and Methods ................................... 31
      3.2.1 Subjects ............................................ 31
      3.2.2 Image Acquisition and Data Processing ....... 31
3.2.3 Nonlinear Mixed Effects Model .................................. 31
3.2.4 Model Formulation ............................................. 32
3.2.5 Model Estimation .................................................. 33
3.2.6 Inference and Predictions ......................................... 33
3.2.7 Regional Analysis of Longitudinal Data Using NLME ......... 33
3.3 Results .................................................................. 34
3.4 Discussion .............................................................. 36
3.5 Conclusions ............................................................. 39

4. PEDIATRIC LONGITUDINAL AUTISM STUDY .................. 42
   4.1 Introduction .......................................................... 42
   4.2 DTI Analysis .......................................................... 42
      4.2.1 Subjects .......................................................... 42
      4.2.2 Image Acquisition and Processing ....................... 43
      4.2.3 Regional Analysis ............................................. 43
   4.3 T1-Weighted Analysis ............................................. 44
      4.3.1 Image Acquisition and Processing ....................... 44
      4.3.2 Statistical Analysis of White Matter Regions .......... 47
   4.4 Discussion .............................................................. 47

5. TWIN STUDY ............................................................ 54
   5.1 Introduction .......................................................... 54
   5.2 Materials and Methods ............................................ 55
      5.2.1 Subjects .......................................................... 55
      5.2.2 Image Acquisition and Data Processing ............... 55
      5.2.3 Statistical Analysis ............................................. 56
   5.3 Results .................................................................. 56
   5.4 Discussion .............................................................. 57

6. MULTIVARIATE NONLINEAR MIXED EFFECTS MODELS ....... 64
   6.1 Introduction .......................................................... 64
   6.2 Extension to Multivariate Analysis .............................. 65
   6.3 Results .................................................................. 65
   6.4 Conclusion .............................................................. 70

7. SUBJECT-SPECIFIC ANALYSIS ........................................ 72
   7.1 Prediction Interval .................................................. 72
   7.2 Evaluation of Individual Scan .................................... 74
   7.3 Prediction of Individual Trajectory .............................. 74
   7.4 Subject-Specific Prediction Interval ............................. 76
   7.5 Conclusion .............................................................. 84

8. DISCUSSION .............................................................. 88
   8.1 Summary of Contributions ......................................... 88
      8.1.1 Statistical Framework ......................................... 89
      8.1.2 Characterization of Longitudinal Changes of MRI Parameters in Multiple Clinical Studies .......... 90
      8.1.3 Subject-Specific Analysis ..................................... 91
   8.2 Limitations ............................................................. 92
LIST OF TABLES

2.1 Comparison of Linear Mixed Effects Model. .......................................................... 22
2.2 Comparison of Two-parameter Monomolecular Model. ........................................ 22
2.3 Three-parameter Monomolecular Model. .............................................................. 23
2.4 Logistic Model. ........................................................................................................ 23
2.5 Gompertz Model. .................................................................................................... 23
3.1 Distribution of Scans Across Different Time Points. .............................................. 31
3.2 Relative Order of Appearance of Myelin from Term to 2 Years. ......................... 38
3.3 Results of Pairwise Testing of all White Matter Regions and All Diffusivity Measures. .......................................................... 39
4.1 Distribution of Scans Across Different Time Points for High Risk Infants .......... 47
4.2 Group Differences in Gompertz Parameters of T1W of White Matter Regions Between HR− and HR+. .......................................................... 49
4.3 Distribution of Scans Across Different Time Points for Healthy Controls (LR−) and High Risk Infants Diagnosed with Autism Spectrum Disorder (HR+). .... 49
4.4 Group Differences in Gompertz Parameters of Longitudinal Trajectories of T1W for White Matter Regions Between LR− and HR+. ......................... 52
5.1 Distribution of Scans Across Different Time Points and Zygosity. ...................... 56
5.2 Gestational Age of Singletons and Twins. .............................................................. 57
5.3 Group Differences in Fractional Anisotropy of White Matter Regions Between Singletons and Twins .......................................................... 59
5.4 Group Differences in Radial Diffusivity of White Matter Regions Between Singletons and Twins. .......................................................... 60
5.5 Group Differences in Axial Diffusivity of White Matter Regions Between Singletons and Twins. .......................................................... 61
7.1 Predicted and Observed Values of FA for Posterior Thalamic Radiation. Neonate and 1 Year Scans Were Used to Predict Values of FA at About 2 Years. .......... 82
7.2 Predicted and Observed Values of FA for Posterior Limb of Internal Capsule. Neonate and 1 Year Scans Were Used to Predict Values of FA at About 2 Years. .......... 82
7.3 Predicted and Observed Values of RD for Posterior Thalamic Radiation. Neonate and 1 Year Scans Were Used to Predict Values of RD at About 2 Years. .......... 83
7.4 Predicted and Observed Values of RD for Posterior Limb of Internal Capsule. Neonate and 1 Year Scans Were Used to Predict Values of RD at About 2 Years. .......... 83
CHAPTER 3

REGIONAL CHARACTERIZATION OF LONGITUDINAL DT-MRI TO STUDY WHITE MATTER MATURATION OF THE EARLY DEVELOPING BRAIN

Regional characterization of longitudinal DT-MRI to study white matter maturation of the early developing brain

Neda Sadeghi a,*, Marcel Prastawa a, P. Thomas Fletcher a, Jason Wolff b, John H. Gilmore c, Guido Gerig a

a Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT 84112, USA
b Carolina Institute for Developmental Disabilities, University of North Carolina, Chapel Hill, NC 27599, USA
c Department of Psychiatry, University of North Carolina, Chapel Hill, NC 27599, USA

A R T I C L E   I N F O
Article History:
Accepted 15 November 2012
Available online 9 December 2012

Keywords:
Longitudinal brain imaging
Early brain development
DTI
Nonlinear mixed effect modeling

A B S T R A C T
The human brain undergoes rapid and dynamic development early in life. Assessment of brain growth patterns relevant to neurological disorders and disease requires a normative population model of growth and variability in order to evaluate deviation from typical development. In this paper, we focus on maturation of brain white matter as shown in diffusion tensor MRI (DT-MRI), measured by fractional anisotropy (FA), mean diffusivity (MD), as well as axial and radial diffusivities (AD, RD). We present a novel methodology to model temporal changes of white matter diffusion from longitudinal DT-MRI data taken at discrete time points. Our proposed framework combines nonlinear modeling of trajectories of individual subjects, population analysis, and testing for regional differences in growth patterns. We first perform deformable mapping of longitudinal DT-MRI of healthy infants imaged at birth, 1 year, and 2 years of age, into a common unblurred atlas. An existing template of labeled white matter regions is registered to this atlas to define anatomical regions of interest. Diffusivity properties of these regions, presented over time, serve as input to the longitudinal characterization of changes. We use non-linear mixed effect (NLME) modeling where temporal change is described by the Gompertz function. The Gompertz growth function uses intuitive parameters related to delay, rate of change, and expected asymptotic value; all descriptive measures which can answer clinical questions related to quantitative analysis of growth patterns. Results suggest that our proposed framework provides descriptive and quantitative information on growth trajectories that can be interpreted by clinicians using natural language terms that describe growth. Statistical analysis of regional differences between anatomical regions which are known to mature differently demonstrates the potential of the proposed method for quantitative assessment of brain growth and differences thereof. This will eventually lead to a prediction of white matter diffusion properties and associated cognitive development at later stages given imaging data at early stages.

Introduction

Improved understanding of typical brain development during infancy, an interval characterized by rapid sculpting, organization and vulnerability to exogenous influences, is of a great importance both for clinical and scientific research. Many neurobehavioral disorders have their origins during neurodevelopment (Gilmour et al., 2010; Huppi, 2008). Establishing a normative model of early brain development is a critical step to understanding the timing and potential mechanisms of atypical development and how intervention might alter such trajectories and improve developmental outcomes (Als et al., 2004; Marsh et al., 2008). Once normative models are available, they can inform research and practice concerning children at risk for neurodevelopmental disorders and may eventually lead to earlier and improved diagnosis and treatment. Longitudinal trajectory-based studies provide a better understanding of human brain development compared to cross-sectional studies (Karnowski-Smith, 2010). In cross-sectional data, calculation of the average trajectory may not be representative for the growth patterns of individual subjects as this approach is inherently insensitive to individual developmental differences and cohort effects (Gogtay et al., 2004). Cross-sectional analysis might falsely report magnitude of changes over time or may fail to detect changes (Casey et al., 2005).

Growth modeling from longitudinal data, on the other hand, makes use of sets of individual temporal trajectories which result in significantly improved models of growth and growth variability, as longitudinal studies can differentiate between cohort and age effects (Diggle et al., 2002). Previous imaging studies of early brain development have substantially contributed to our current understanding of brain development. Some of the studies considered size or shape differences (Huppi, 2008; Knickmeyer et al., 2008; Xu et al., 2008; Xue et al., 2007), others have looked at changes of contrast in MRI (Sadeghi et al., 2010) or
diffusion parameters in DTI (Gao et al., 2009; Geng et al., 2012; Hermoye et al., 2006; Huppi et al., 1998; Mukherjee et al., 2002; Sadeghi et al., 2012). However, most of these studies are based on cross-sectional data or children older than 2 years (Dubois et al., 2008; Faria et al., 2010; Gao et al., 2009; Hermoye et al., 2006; Mukherjee et al., 2002). In this study we focus on developing longitudinal models spanning birth to about two years of age. The models are based on the parameters obtained from diffusion tensor imaging (DTI). DTI-derived diffusivity parameters provide relevant information about the maturation of the underlying tissue as they assess water content (Huppi, 2008). These measurements are a possible reflection of axonal density and/or degree of myelination (Neill et al., 1998; Song et al., 2002) which correlate with cognitive functions (Dubois et al., 2006) and early developmental outcomes (Als et al., 2004; Ment et al., 2009; Wolff et al., 2012). In this study we focus on fractional anisotropy (FA), mean diffusivity (MD), radial (RD) and axial diffusivity (AD) to explain brain maturation and to gain a better understanding of white matter development. Driven by earlier findings that myelination follows a nonlinear spatio-temporal pattern (Dubois et al., 2008), our goal is to capture these changes in terms of the parameters of the Compertz function which provides an intuitive parameterization representing delay, growth, and asymptotic values for each region.

In contrast to previous studies, we use an explicit growth function (the Compertz function) and a nonlinear mixed effects modeling scheme (Pinheiro and Bates, 2000). In a nonlinear mixed effects model, the diffusion parameters are modeled in a hierarchical fashion, with fixed-effect representing the overall population trend, and random effect associated with each individual. Nonlinear mixed effect models are suited for longitudinal data where each subject has repeated scans with the possibility of missing data points and uneven spacing between scans of all the individuals in the group. Unlike most previous studies of early brain development, we make use of longitudinal imaging where each subject is imaged repeatedly over the first few years of life. This enables a more accurate characterization of developmental pattern (Giedd et al., 1999). Nonlinear mixed effect model provides a direct way of estimating individual trajectories along with longitudinally derived typical developmental curves as illustrated in Fig. 2. This leads to the characterization of a normative model for healthy developmental patterns and estimation of personalized, individual trajectories of growth, which is a property that will be desirable for comparison and diagnostic assessment of individual subjects.

We apply our analysis framework to a set of white matter regions that are known to have different patterns of growth to establish normative developmental patterns for each region. Quantitative analysis of diffusion changes in these regions provide further insight into brain maturation process and will enable prediction of subject-specific growth trajectory with the potential of detecting pathological deviation related to brain disorders.

Materials and methods

Subjects

This study was approved by the Institutional Review Board of the University of North Carolina School of Medicine. Children analyzed in this study are controls in an ongoing longitudinal study of early brain development in high risk children (Geng et al., 2012). A total of 26 control subjects were selected for this study. Scans of these subjects were obtained at around two weeks, 1 year and 2 years. Four of the subjects had sub-optimal scans at 1 year that were removed, but their scans for other time points were kept. In total, we used 59 datasets, the temporal distribution of scan data is shown in Table 1. To ensure maximal success rate of scanning, all subjects were fed, swaddled and fitted with ear protection. All subjects were scanned without sedation during their natural sleep.

<table>
<thead>
<tr>
<th>Available scans</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neonate scan only</td>
<td>2</td>
</tr>
<tr>
<td>1 year scan only</td>
<td>0</td>
</tr>
<tr>
<td>2 year scan only</td>
<td>0</td>
</tr>
<tr>
<td>Neonate + 1 year scan</td>
<td>10</td>
</tr>
<tr>
<td>Neonate + 2 year scan</td>
<td>2</td>
</tr>
<tr>
<td>1 year + 2 year scan</td>
<td>3</td>
</tr>
<tr>
<td>Neonate + 1 year + 2 year scan</td>
<td>9</td>
</tr>
</tbody>
</table>

### Image acquisition and data processing

All images were acquired using a 3 T Allegra head-only MR system using a single shot echo-planar spin echo diffusion tensor imaging sequence with the following parameters: TR = 5200 ms, TE = 73 ms, slice thickness of 2 mm and in-plane resolution of 2 × 2 mm². One image without diffusion gradients (b = 0) along with 6 gradient directions with a b-value of 1000 mm²/s were acquired. The sequence was repeated 5 times for improved single-to-noise ratio. All DWIs were checked and corrected for motion artifacts using the DTIchecker tool. Tensor maps were calculated for each DTI scan using weighted least squares tensor estimation on the images that have been averaged over sequence repeats (Salvador et al., 2005). T2-weighted structural images were obtained using turbo spin echo sequence with TR = 7 s, TE = 15 and 90 ms, slice thickness of 1.95 mm and in-plane resolution of 1.25 × 1.25 mm². T2W and baseline DWI of all the subjects' scans were skull stripped using Brain Extraction Tool (BET) (Smith, 2002).

Due to significant contrast changes in early brain development, we utilized two registration frameworks: one for intra-subject and inter-modality registration, and the other for inter-subject registration. For intra-subject registration, we applied the unbiased atlas building framework of Joshi et al. (2004) based on the Large Deformation Diffeomorphic Metric Mapping (LDDMM) (Miller et al., 2002) to the set of T2W images of scans at year 1 to obtain spatial mappings between all subjects through the estimated atlas coordinate system. Intra-subject registration was performed by IRTK software, using affine and nonlinear registration method of Rueckert et al. (1999) using normalized mutual information as the image match metric (Studholme et al., 1999) that appears robust to changing contrast properties in early brain development. All time points of each subject are registered to the unbiased atlas via linear and non-linear transformations, first by mapping these images to the year 1 scan and then cascading the two transformations for a mapping to the atlas. Details on the registration methods and parameters are summarized in Appendix A. The tensors are registered to the atlas using transformations obtained by registering the DTI baseline (0) images to T2W images. Tensors are resampled using finite strain reorientation and Riemannian interpolation (Alexander et al., 2001; Fletcher and Joshi, 2007; Pennec et al., 2006). After all the images are transformed into the atlas space, the tensors are averaged using the log-Euclidean method to produce a tensor atlas (Arsigny et al., 2006). In this study, we extract the mean, axial, radial diffusivity, and fractional anisotropy features from the registered tensors, MD = \( \mu_{\text{axial}}, AD = \lambda_1, \) RD = \( \lambda_2 + \lambda_3, \) and FA = \( \sqrt{\lambda_2 \cdot \lambda_3}/\sqrt{\lambda_1 \cdot \lambda_2 \cdot \lambda_3}, \) where \( \lambda_1 \) are the eigenvalues of the tensor sorted from largest to smallest. Fig. 1 shows an overview of our method and analysis workflow.

### Nonlinear mixed effects model

In this subsection, we describe the nonlinear mixed effects model used to analyze the longitudinal DTI data. Compared to a nonlinear model...
least squares (NLS) method, a nonlinear mixed effects (NLME) model does not assume that the sample data points are independent and identically distributed, rather it assumes that there is correlation across repeated measurements. Also, the average trend estimated based on the mixed effect model is an average of individual trajectories rather than a least squares fit to the individual data points. This results in better representation of trajectories in the population as illustrated in Fig. 2.

Model formulation

In the mixed effects model, the observed data is a combination of fixed effects which are parameters associated with the entire population or a sub-population, and random effects which are parameters associated to an individual. In the nonlinear mixed effect models, some or all the parameters appear nonlinearly in the model. We use the NLME model proposed by Lindstrom and Bates (1990) where each individual's observation is modeled as:

$$ y_{ij} = f(i, t_0) + e_i $$

where $i$ indexes the individual subjects and $j$ indexes the time points, $M$ is the number of individuals, $e_i$ is the number of observations on the $i$th individual, $f(i, t_0)$ is a nonlinear function of the covariate vector (time) $t_0$ and parameter vector $\phi_i$ and $e_i \sim N(0, \sigma^2)$ is an i.i.d. error

---

**Fig. 1.** Overview of the proposed longitudinal DTI region based analysis.
term. The parameter vector can vary among individuals by writing \( \Phi_i \) as

\[
\Phi_i = A_i \beta + B_i h \sim \mathcal{N}(0, \Psi)
\]

(2)

where \( \beta \) is a \( p \)-vector of fixed effects, and \( h \) is a \( q \)-vector of random effects associated with individual \( i \) with variance-covariance \( \Psi \). \( A_i \) and \( B_i \) are identity matrices for our study.

The function \( f \) can be any nonlinear function. Since early brain development is characterized by rapid initial development which slows down in later years, it is preferable to use growth functions which reflect these properties. One such growth function is the Gompertz function which can be written as:

\[
y = \text{asymptote} \exp(-\text{delay}) \exp(-\text{speed} \cdot t).
\]

(3)

The effects of varying the three parameters asymptote, delay, and speed of the Gompertz function are shown in Fig. 3, for a function that decreases as time progresses.

To use the Gompertz function in the nonlinear mixed effect model, we apply the following formulation where the Gompertz function is parameterized as \( y = f(\phi, t) = \phi_1 \exp\left(-\phi_2 \phi_3^t\right) \), where \( \phi_1 \) denotes asymptote, \( \phi_2 \) is delay, and \( \phi_3 \) is \exp(-speed). Combining the nonlinear mixed effect model with the Gompertz function, each observation can be represented as follows:

\[
y_{ij} = f(\phi_i, t_j) + \epsilon_{ij} = \phi_{i1} \exp\left(-\phi_{i2} \phi_{i3}^{t_j}\right) + \epsilon_{ij}
\]

(4)

where the mixed effects are \( \phi_i = [\phi_{i1}, \phi_{i2}, \phi_{i3}]^T = \beta + b_i \), the fixed effects are \( \beta = [\beta_1, \beta_2, \beta_3]^T \), and the random effects for each subject \( i \) are \( b_i = [b_{i1}, b_{i2}, 0]^T \). We set one of the random effects to zero to reduce the number of random effects in the model. As we only have a maximum of three time points per subject, including an additional random effect may cause the matrix \( \Psi \) to be rank-deficient (singular) and thus create problems in the estimation of the parameters.

**Model estimation**

Different methods have been proposed to estimate the parameters as shown in Eq. (4). Since random effects are unobserved quantities, we use the marginal density of responses \( y \) to obtain the parameters of the nonlinear mixed effects model. The following maximum likelihood estimation is performed to obtain the parameters of Eq. (4):

\[
y_i \sim \mathcal{N}\left(y_i \mid \phi_i, \Psi\right) = \mathcal{N}\left(y_i \mid \beta + b_i, \Psi, \sigma^2\right) p(h_i) dh_i
\]

(5)

Due to nonlinearity presented in the random effects of function \( f \), there is generally no closed form solution to the integral. Here, we use the estimation method proposed by Lindstrom and Bates (1990) using the nlme package (Pinheiro et al., 2012) in R\(^2\) to obtain the model parameters. This algorithm iterates between two steps: a penalized nonlinear least square step and a linear mixed effects step until convergence.

**Inference and predictions**

Under the linear mixed effects approximation, the distribution of maximum likelihood estimators \( \hat{\beta} \) of the fixed effect is:

\[
\hat{\beta} \sim N\left(\beta, \sigma^2 \left(\sum_{i=1}^{n} \chi_i \chi_i^T\right)^{-1}\right)
\]

(6)

where \( \chi_i = I + 2z_i \Delta^{-1} \Delta^{-1} X_i \), \( X_i = \frac{\partial y_i}{\partial \beta}|_{\beta=\hat{\beta}} \), \( \Delta = \frac{\partial^2 y_i}{\partial \beta \partial \beta^T}|_{\beta=\hat{\beta}} \), and \( \Delta \) is the precision factor such that \( \Psi^{-1} = \sigma^{-2} \Delta^{-1} \Delta \) (Pinheiro and Bates, 2000).

Knowing fixed parameters \( \hat{\beta} \) and its sampling distribution, it is straightforward to conduct hypothesis testing among different regions or between healthy and/or at-risk populations. We can also obtain individual growth trajectories based on the estimated random effects for each individual. For example, the individual response for subject \( i \) is \( \hat{y}_i = f(\beta + b_i, t) \), and the population growth trajectory is estimated when random effects are set to their mean value, \( \bar{G} \), resulting in \( \hat{y}_i = f(\bar{G}, t) \).

**Regional analysis of longitudinal data using NLME**

We use the nonlinear mixed effects model to model the longitudinal DTI data within anatomical regions and perform hypothesis testing between trajectories of these regions. Maps of these anatomical regions were developed and disseminated by Mori et al. (2008), and mapped to our unbiased atlas via linear followed by nonlinear B-spline registration (Rueckert et al., 1999). We select 13 anatomical regions in the atlas space as shown in Fig. 4. In this study, left and right regions of anatomical locations are combined, giving a total of

3 http://n-project.org.
eight regions. Future studies on lateralization of growth differences will analyze left and right regions separately. The labeling of regions in the atlas space allows automatic partitioning of each subject's scans into the different anatomical regions. We then estimate growth trajectories for these regions using the NLME model (Lindstrom and Bates, 1990) described previously. The mixed parameters are the asymptote \( \phi_1 \), delay \( \phi_2 \) and speed \( \phi_3 \) of the Gompertz function for each region, which requires a slight modification to Eq. (4) to account for regions:

\[
y_{ij} = f(\phi_1, \phi_2) + \epsilon_{ij} = \phi_1 \exp\left( -\phi_2 \phi_3 t_i \right) + \epsilon_{ij}
\]  

(7)

We then conduct hypothesis testing between pairs of regions to determine modes of longitudinal changes in terms of the Gompertz growth parameters. With \( N \) number of regions, we perform \( \frac{N(N-1)}{2} \) pairwise fitting of nonlinear mixed effect modeling. The significant parameters are determined through t-tests, corrected for multiple comparisons by Bonferroni correction. The parameters that are found to be significant between two pairs of regions can be interpreted as the distinguishing feature between the longitudinal trajectories of these regions.

Results

We applied our framework to longitudinal pediatric DTI data of 26 subjects. In total, we selected 13 regions in the unbiased atlas as shown in Fig. 4. The regions are as follows: anterior limb of internal capsule (right and left; ALIC), posterior limb of internal capsule (right and left; PLIC), genu, body of corpus callosum (BCC), splenium (Sp), external capsule (right and left; ExCap), retrolenticular part of internal capsule (right and left; RLC), and posterior thalamic radiation which includes optic radiation (right and left; PTR). The right and left of each anatomical region were combined giving a total of eight regions. Fig. 5 plots the average FA, MD, RD, and AD of each region for each subject. In all the regions, FA increases with age, whereas MD, RD and AD decrease with age. Interestingly, each region develops in a distinctly different temporal pattern.

![White matter anatomical structures](image)

**Fig. 4.** White matter anatomical labels that are used for regional analysis. Labels are overlaid on the FA (Fractional Anisotropy) map of the reference space that is the population atlas.
Fig. 5. Plots of diffusivity measures (FA, MD, AD and RD) versus age, shown for 26 control subjects and eight regions. Colors indicate different regions (purple: A1C, light green: ExCap, brown: Genu, blue: PLIC, dark green: PTR, red: BCC, yellow: Sp, orange: ICC), solid lines connect the mean of each region. In all the regions, FA increases with age, whereas MD, RD and AD decrease with age. Interestingly, each region develops a distinctly different temporal pattern.

Paired t-tests of growth trajectories were performed for all combination of pairs of regions for all the diffusion parameters. The results of all pairwise comparisons can be found in Table 3 in Appendix B. Differences in parameters $\beta_1$ and $\beta_2$ were significant between most pairwise comparisons among diffusion parameters, whereas $\beta_3$ was only significant in a few regions: genu, splenium, and body of corpus callosum, and mostly when considering the RD or MD measurements. Genu was the only structure that was significantly different than all the other regions in the $\beta_3$ parameter of RD and MD. This region decreased in MD and RD at a slower rate compared to all the other regions. We didn't find any pattern that was consistent among different parameters and different measurements since each parameter measures a different aspect of growth. Interestingly, we noticed some pairwise comparisons with significant differences in $\beta_1$ parameter between AD and RD trajectories, but no differences in MD (A1C vs. PLIC, Genu vs. ExCap). This happens when reverse temporal patterns are seen for AD and RD, suggesting that analysis of AD and RD may reveal much better insight into maturation than MD alone.

In this section, we focus on PLIC/A1C, body of corpus callosum (BCC), and splenium comparisons as examples of commissural and projection fibers. These regions are known to have a distinctive maturation pattern and axonal density. The PLIC is one of the structures

Fig. 6. Population and individual growth trajectories for PLIC and A1C regions. Thicker curves illustrate the average growth trajectories, and individual trajectories are shown via the red and blue functions of individual subjects for A1C and PLIC, respectively. Compare parameters with statistically significant differences are: FA: $\beta_1$, $\beta_2$, MD: $\beta_3$, RD: $\beta_1$, (AN), AD: $\beta_2$, where * denotes $p<0.05$, ** denotes $p<0.01$ and where $\beta_1$, $\beta_2$ and $\beta_3$ represent asymptote, delay and speed.
that shows early myelination, while ALIC shows later maturation compared to PLIC as is shown in higher FA, and lower RD and MD. Fig. 6 shows the population and individual trajectories of FA, MD, RD, and AD as modeled by Nonlinear Mixed Effect for ALIC/PLIC. As expected, the PLIC shows a higher FA compared to ALIC at birth mainly explained by lower RD. After about 800 days both regions have the same MD and similar FA and RD values. However, the ALIC shows a higher AD compared to PLIC, possibly indicating a different structuring of this tract region. The delay parameter of the Compertz function $\hat{\beta}_3$ was significantly different between ALIC and PLIC for FA, MD, and RD measurements, an indication of later development of ALIC compared to PLIC. Also, the asymptote $\hat{\beta}_3$ was significantly different for FA, RD, and AD.

The body of the corpus callosum (BCC) and splenium (Sp) are known to have very limited myelination at birth but higher axonal density compared to ALIC and PLIC, and the splenium shows earlier myelination compared to BCC (Rutherford, 2002). Fig. 7 shows population and individual growth trajectories for the body of the corpus callosum and splenium. The splenium shows higher FA at birth and also throughout the first two years, while RD is about same at birth, but diverges at two years. Reverse patterns are seen for AD and RD at about two years, which causes RD to be about the same. All three parameters of the Compertz function for RD were significantly different between BCC and Splenium, suggesting that RD may capture early maturation patterns more sensitively than the other measures. The asymptote parameter was significantly different among all the measurements between these two regions.

Fig. 8 shows FA, RD, and AD of PLIC (shown in blue) compared to the other three regions ALIC, BCC, and Sp (shown in red). In this figure, solid lines are the average estimated growth trajectories for each region, the shaded regions are the 95% confidence interval of these average curves. Monte Carlo simulation was used to generate 1000 curves based on the approximate distribution of the maximum likelihood estimates of fixed effects. The 95% range of these curves are calculated pointwise to obtain the confidence interval. The dashed lines show the 95% prediction interval which is also calculated based on the Monte Carlo simulation of 1000 curves based on the approximate distribution of both fixed effects and random effects. The splenium shows a high RD at birth relative to PLIC, by about 800 days however, both regions have approximately the same RD value as shown in Fig. 8. The splenium has very limited myelination at birth, while the PLIC is known to have a higher level of myelination at this time of development. These facts are evident in the difference in RD at birth between splenium and PLIC. At age two, however, the splenium shows approximately the same RD value, indicating that it catches up with PLIC.

The values of Compertz parameters for all the regions and all diffusivity measures are shown in Fig. 9. Each region shows a distinct pattern of development as is depicted by the $\hat{\beta}_3$, $\hat{\beta}_4$, and $\hat{\beta}_5$ parameters of Compertz function. As indicated in the section ‘Model formulation’ the parameters $\hat{\beta}_3$, $\hat{\beta}_4$, and $\hat{\beta}_5$ represent asymptote, delay and speed, respectively. When $\hat{\beta}_3$, $\hat{\beta}_4 > \hat{\beta}_5$, the expected value of diffusion parameters for region A is higher than region B at year 2. When $\hat{\beta}_3$, $\hat{\beta}_4 > \hat{\beta}_5$, region $\hat{\beta}_5$ matures earlier compared to $\hat{\beta}_4$. The scenario $\hat{\beta}_3 > \hat{\beta}_4 > \hat{\beta}_5$ indicates accelerated growth for $\hat{\beta}_3$ compared to $\hat{\beta}_4$. Note that the delay parameter is negative for RD, AD, and MD measurements as these values decrease during early brain development, where as the delay parameter is positive for FA as fractional anisotropy increases during this time period.

Discussion

Assessment of brain growth patterns in these regions reveals a nonlinear pattern of maturation with considerable regional variation as shown in previous studies (Hermoye et al., 2005; Mukherjee et al., 2001; Schneider et al., 2004). In agreement with previous studies, increased FA and decreased MD, AD, RD were observed within all the white matter regions during this period (Forbes et al., 2002; Mukherjee et al., 2001; Schneider et al., 2004; Zhang et al., 2005). This longitudinal pediatric study supports a rapid change during the first 12 month followed by slower maturation during the second year similar to previous studies (Geng et al., 2012; Hermoye et al., 2006). Our study, in addition to supporting earlier cross-sectional reports on negative correlation between age and diffusion parameters, provides greater statistical power to examine nonlinear pattern of maturation in various white matter regions.

Beyond the analysis of FA and MD measurements, in this study we included RD and AD analysis of these white matter regions. The regional comparisons of white matter regions indicates that
individual AD and RD carry important information which may not be found in the MD diffusivity measures. The relationship of AD/RD and FA is complex and nonlinear, but our data suggest that modeling FA, AD, and RD as time trajectories provides more information than only FA as illustrated in Figs. 6 and 7.

For example, FA of splenium and PLIC are approximately the same values at birth, yet we know that the splenium is not myelinated at birth, and we see the significant differences of RD between these regions. The high FA value of the splenium at birth may be due to its high density of axons. This discussion of FA for PLIC and splenium clearly reflects that FA is not necessarily a good indicator for the degree of myelination and may be greatly influenced by axonal density particular to this developmental interval (LaMantia and Rakic, 1990). In contrast, the similarity of FA trajectories for PLIC and splenium, for which we see very different AD and RD patterns and thus different tensor shapes, illustrates that interpretation of FA with respect to myelination and structural integrity is difficult, and that the additional AD and RD measures provide richer information.

Modeling the nonlinear growth changes of white matter by the Gompertz function and inclusion of AD and RD to the analysis provides a more detailed and comprehensive picture of the changes within these white matter regions. Compared to previous studies of linear fitting with logarithm of age (Chen et al., 2011; Faria et al., 2010; Lobel et al., 2009) we fit the nonlinear growth curves (Gompertz function) to the diffusion data and actual age, this enables the parameterization of the trajectories in terms of asymptote, delay and speed and models nonlinear temporal changes with improved accuracy. Based on our finding, the delay parameter of the Gompertz function, \( \beta_2 \) of RD seems to be closest related to myelination process if we compare results to what is known from the literature. Looking at RD and \( \beta_2 \) delay parameters of the Gompertz function as is shown in Fig. 5, we see a good correspondence with previous radiology findings, such as in Rutherford (2002). In fact, RD has been considered to be in correspondence with histological changes in demyelination (Song et al., 2002). Table 2 compares our findings versus existing knowledge from radiology literature, which indicates development of PLIC prior to AUL, and splenium prior to genu which is also consistent with previous histological findings (Brody et al., 1987; Kinsey et al., 1988).

Our framework is designed not only to provide qualitative comparisons, but to give researchers and clinicians quantitative parameters and a statistical testing scheme. Moreover, the method includes modeling of growth trajectories of individuals, resulting in personalized profiles. This property will be crucial for efforts to improve prediction and diagnosis for individuals, as well as partitioning groups of subjects according to subtypes and subtle variations in early developmental trajectories. Models which assume invariance or linearity between neurobehavioral markers are apt to miss crucial shifts in development (Shaw et al., 2006; Thomas et al., 2009). The ability of the present framework to capture the dynamic properties of inter- and intra-individual development has the potential to substantially improve clinical applications of developmental neuroimaging.

There are some limitations to our proposed framework. Our analysis depends on accurate image registration among all the subjects and time points. Early brain development is characterized by a rapid change of contrast and size of the brain, which makes registration a challenging task. However, in this study we decided to use ROI defined regions

![Image of graphs showing PLIC (blue) vs. ALIC (red), PLIC (blue) vs. Sp (red), PLIC (blue) vs. BCC (red)]
Table 2

<table>
<thead>
<tr>
<th>Distribution of myelin as seen in T1W and T2W by Rutherford</th>
<th>Estimated based on RD delay parameter $f_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLIC and optic radiation</td>
<td>PLIC, PTR and ExCap</td>
</tr>
<tr>
<td>AUC</td>
<td>AUC and BCC</td>
</tr>
<tr>
<td>Not available</td>
<td>RUC</td>
</tr>
<tr>
<td>Splenium</td>
<td>Splenium</td>
</tr>
<tr>
<td>Genu</td>
<td>Genu</td>
</tr>
</tbody>
</table>

which we expect to be more robust to misregistration compared to voxel-based analysis, and these regions are located more interiorly where we expect less registration problems. Nonetheless, improved spatial registration will potentially improve the accuracy of the model. Another limitation is that the statistical analysis is based on the log-likelihood of nonlinear mixed effects modeling, which does not have a closed form solution. We have used a linear mixed effect approximation, so greater care should be taken when doing hypothesis testing with the estimated parameters.
In the future, we plan to extend our method to tract-based regions with modeling along the tract changes. We also plan to extend the model to multivariate growth function similar to (Xu et al., 2008) and include a much larger set of regions for analysis.

Conclusions

We have presented a framework for the processing of longitudinal images in order to characterize longitudinal development of white matter regions at both the individual and group level. By utilizing nonlinear mixed effects modeling, we jointly estimate the population trajectory along with each individual trajectories. Compertz parameterization of diffusion changes provides an intuitive parameterization of growth trajectory in terms of asymptote, delay and speed. This provides a description of longitudinal changes with potential for detecting deviations from a typical growth trajectory sensitive to multiple neurodevelopmental phenomena. We have also presented a method for making inference about regional differences in diffusion properties known to vary by microstructural properties and developmental course (Dubois et al., 2009; Kinney et al., 1998; LaMantia and Rakic, 1992; Lebel and Beaulieu, 2011). This is in contrast to standard modeling and analysis of testing for group or regional differences as it reveals the type, timing, and nature of differences. The proposed analysis can be extended to an arbitrary number of regions, and applied to other measurement such as structural MRI.

As discussed in the previous section, the present study clearly illustrate that studying FA alone as an indicator of white matter maturation or integrity insufficiently characterizes structural properties of white matter and may produce misleading results as regions with very different axonal density and differing degrees of myelination may show similar FA values. We suggest that in addition to FA, studies should include statistical analysis of AD and RD, which provide important additional information to better explain FA measures. In regard to early maturation, we demonstrate that the radial diffusivity (RD) measure and the delay parameter \( \beta_d \) of the Compertz function seem to be the best combination to describe early brain maturation. We will further explore this in applying our framework to DTI of infants with developmental delay and myelination storage disorders such as Krabbe's disease.

Acknowledgments

Supported by NIH grants: R01 MH070890 (JHG, GC), Conte Center MH064065 (JHG, GC), National Alliance for Medical Image Computing (NA-MIC) U54 EB005149 (GG) and the Utah Science Technology and Research (USTAR) initiative at the University of Utah.

Appendix A. Summary of registration parameters

Intra-subject and inter-modality registration

We use the IRTK software (Rueckert et al., 1999) to perform intra-subject and inter-modality registration. The registration method is a multi-scale approach using B-spline transformation, where we use the normalized mutual information image match metric. We use three different scales and discretize the image intensity histograms into 64 bins. In this study, the B-spline transforms are parameterized using \( 14 \times 14 \times 14 \) control points.

Inter-subject registration

We construct an unbiased atlas (Joshi et al., 2004) and the associated inter-subject registration using the Large Deformation Diffeomorphic Metric Mapping (LDDMM) (Miller et al., 2002) that minimizes the following objective function:

\[
\text{arg min}_{\phi} \sum_i \left\| l_i - L_i \cdot \phi_0 \right\|_2^2 + \sum_{f=1}^F \int_{\Omega} \left\| \phi_i \right\|_2^2
\]

where \( \text{I} \) is the image atlas, \( l_i \) is the image of subject \( i \), \( \phi_i \) is the mapping relating subject \( i \) to the atlas that is parameterized using the velocity \( v_i \). Regularity of the mapping \( \phi_i \) is enforced by minimizing

\[
\left\| v_i \right\|_2^2 = \left\langle v_i, v_i \right\rangle = v_i^T \nabla v_i + \beta \nabla v_i + \gamma \nabla \nabla v_i
\]

Table 3

Results of pairwise testing of all white matter regions and all diffusivity measures. Compertz parameters with significant differences are denoted by * for \( p < .05 \) and ** for \( p < .01 \). Non-significant parameters are indicated by 'ns'.

<table>
<thead>
<tr>
<th>Region</th>
<th>FA</th>
<th>MD</th>
<th>RD</th>
<th>AD</th>
<th>Diffusivity</th>
<th>PTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALC</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>MD</td>
<td>NA</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>RD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>AD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>FIC</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>MD</td>
<td>NA</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>RD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>AD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>Genu</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>MD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>RD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>AD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>ICC</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>MD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>RD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>AD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>Sp</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>MD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>RD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>AD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>ExCap</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>MD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>RD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>AD</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
<tr>
<td>Rlic</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
<td>1+</td>
</tr>
</tbody>
</table>