Industry allocated project number

SATI	CFPA	SAAPPA/SASPA	DFTS	Winetech
tarryn@satgi.co.za Tel: 021 872-1438	inmaak@mweb.co.za Tel: 021 872-1501	theresa@hortgro.co.za Tel: 021 882-8470	dappies@dtd.co.za Tel: 021 870 2900	andraga@winetech.co.za Tel: 021 807 3387
				X

FINAL REPORT
2014

Programme & Project Leader Information

	Research Organisation Programme leader	Project leader
Title, initials, surname	Prof JT Burger	Dr HJ Maree
Present position	Professor	Researcher
Address	Dept. of Genetics Stellenbosch University 7600 Stellenbosch	ARC-BTP / Dept. of Genetics Stellenbosch University 7600 Stellenbosch
Tel. / Cell no.	021-8085858	021-8085885
Fax	021-8085833	021-8085833
E-mail	jtb@sun.ac.za	hjmaree@sun.ac.za

Project Information

Research Organisation Project number	GenUS 10-01
Project title	A metagenomic sequencing approach to elucidate Shiraz disease etiology.

Fruit kind(s)	Wine Grapes		
Start date (mm/yyyy)	01/01/2011	**End date** (mm/yyyy)	31/12/2013

Project keywords	Shiraz disease

Approved by Research Organisation Programme leader (tick box)

This document is confidential and any unauthorised disclosure is prohibited.

THIS REPORT MUST INCLUDE INFORMATION FROM THE **ENTIRE** PROJECT

Executive Summary

Several projects have attempted to determine the causative/etiological agent(s) associated with the important virus disease complexes found in South African vineyards, with mixed success. The etiology of Shiraz disease has proven to be a particularly elusive problem and this project aimed to elucidate the complex viral disease etiology associated with Shiraz disease (SD) with a metagenomic sequencing approach using the latest in next-generation sequencing (NGS) technology.

The virus profiles of individual vines were constructed by sequencing dsRNA extracted from phloem scrapings. The profiles of SD affected vines were compared to non-SD (GLD) control vines. Diagnostic RT-PCR assays were developed to validate novel viruses detected by the NGS.

The project commenced in 2011. Initially the extraction of high quality dsRNA from grapevine phloem scrapings was optimised. The existing protocols for the extraction of dsRNA were adjusted and further optimised to ensure consistency and the high quality required for sequencing. Sixty-nine samples (46-SD, 23-controls) were collected from several wine farms from different geographical regions (Vredendal, Stellenbosch, Paarl, Franschhoek and Rawsonville). Fifteen samples that yielded high quality dsRNA were used to construct cDNA libraries for sequencing. These samples were sequenced and the NGS data analysed.

Bioinformatic analysis indicated a high level of virus complexity in SD vines compared to the control samples. The viruses identified were the commonly detected viruses such as: GLRaV-2, GLRaV-3 (several variants), GFkV, GVA (several variants), GVB, GVE, GVF as well as several mycoviruses and four different viroids. A new endornavirus, that we named grapevine endophyte endorna virus (GEEV), was also discovered.

The association of a specific virus profile with SD could not be established. However, GLRaV-3 was always present in SD samples in combination with mostly GVA or sometimes GVE. An interesting observation that GVA from genetic variant group II was never observed in GLD control plants, was also made. No new viruses were discovered that were exclusively associated with the SD samples.

It would seem from the NGS results that the viral composition of SD is complex with yet to be understood interactions between known viruses. There is also the possibility of other factors such as host genetics or additional pathogens (bacterial, fungal or even phytoplasma) that could influence symptom development.

Problem identification and objectives

Viral diseases have a detrimental affect on both the quality and yield of grapes as well as the longevity of infected vineyards. The major viral disease in South African vineyards is Leafroll, however, new diseases, such as Shiraz disease (SD) and Shiraz decline are becoming more prevalent. The etiological agents of these diseases are unclear at present. The aim of this project is to elucidate the complex **viral disease etiology** associated with SD with a metagenomic sequencing approach using the latest technology in next generation sequencing.

Objectives of the project:

- To identify the viruses (incl. variants) associated with SD through deep sequencing of diseased and healthy plants.

This document is confidential and any unauthorised disclosure is prohibited.

- To design accurate diagnostic assays for SD and confirm association of viruses (incl. variants) in larger survey.
- To determine the biological contributions of individual viruses to the etiology of SD.
- To plan effective disease control strategies.

Workplan (materials and methods)

Shiraz diseased vines along with control samples were collected and double-stranded RNA (dsRNA) extracted from individual vines. DsRNA was used for the construction of Illumina sequencing libraries and sequenced at the ARC Biotechnology Platform.
Several sequencing runs were planned with the number of samples per run depending on the amount of multiplexing that the service provider and the cost. Three sequencing runs were performed. Data analysis of these NGS datasets were analysed with several software packages both free and commercial. The obtained results for the individual samples were compared to identify a possible correlation between specific viruses or their variants and the expression of SD. Specific primers binding to the novel viruses that were discovered were designed to be used in RT-PCR diagnostic assays. These included grapevine endophyte endorna virus (GEEV) and a new divergent variant of GLRaV-3 also detected in project GenUS11-3.

Results and discussion

Raw data clean-up:
NGS read datasets were trimmed and filtered for downstream analysis. Read lengths of complete datasets were hard trimmed at the position where the average quality score reached 28. The first 9 bases were also removed due to per base nucleotide imbalance. The datasets were then filtered for quality with the variable parameters such that a minimum of 96% of the read had a quality score higher than 20. Figure 1, depicts the improvement on the data quality through this approach. High quality datasets were further filtered for host sequences (genome, chloroplast and mitochondria) before being used in *de novo* assemblies and read-mapping analyses with CLC genomics workbench.

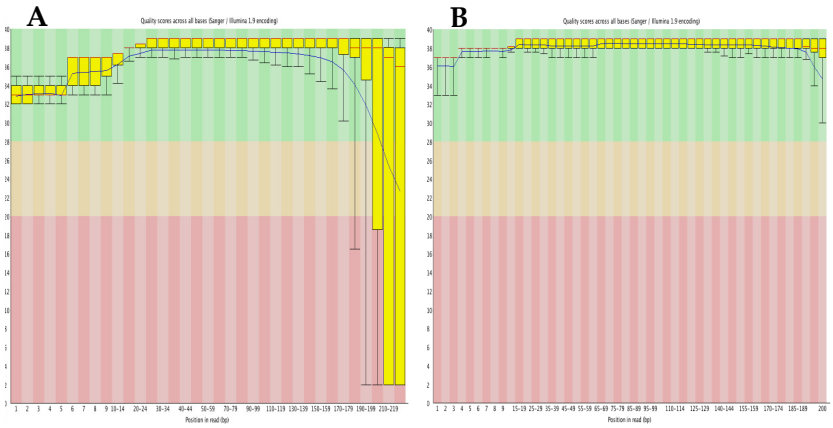


Figure 1: The "per base sequence quality score" Box whiskerplot output graphs by FastQC of the same data set before (A) and after (B) data clean-up. The central red line is the median value, the yellow box represents the inter-quartile range (25-75%), the upper and lower whiskers represent the 10% and 90% points and the blue line represents the mean quality. The Y-axis is the quality score and the X-axis the nucleotide base position. In panel A, an Illumina MiSeq dataset with wide whiskers and inter-quartile range towards the 3' end of the read data. Panel B show the same dataset on which a hard trim of the first 9 nt and last 11 nt was performed followed by quality filtering.

De novo assembly:

This document is confidential and any unauthorised disclosure is prohibited.

De novo assemblies were performed with a range of free bioinformatic tools but after careful considerations and due to the performance of CLC genomics workbench the analysis was moved to this commercial package. Contigs generated were identified through BLAST analysis on GenBank. The analysis of *de novo* assembled contigs revealed the presence of several viruses known to infect grapevine and commonly found in South African vineyards. Fungal viruses were only classified to family level by amino acid similarity due to the lack of sufficient information in sequence databases. Fourty-six different mycoviruses were identified belonging to at least 6 families. Even though, the complexity of virus infection was greater in SD plants than in the control plants, no association with specific viruses can be confirmed. Additionally, four different viroids (GYSVd-1, GYSVd-2, AGVd, and HSVd) were, for the first time, identified in South African vineyards. GLRaV-3 was identified in all samples and GVA predominantly in the SD samples. Additionally, a new genetic variant of GLRaV-3 (also in project GenUS 11-3) and a new endorna virus that we named grapevine endophyte endorna virus (GEEV) was identified.

Read-mapping:

The read-mapping analysis was performed on CLC genomics workbench since it provided to be the most convenient and user-friendly interface. In this focussed approach 118 accessions from GenBank was used as reference sequences to which the NGS data was mapped. These accessions included common, minor and rare grapevine viruses as well as, viroids, grapevine mycoviruses and other mycoviruses (see table 1). Mapping parameters for reads were set at a minimum of 95% similarity for a read to be mapped. Reads that could bind to multiple sequences were allowed to map randomly. The result of the read-mapping can be seen in table 2. “Hits” were classified into four categories according to the percentage of the reference genome that was covered see bottom of table 2.

The read-mapping analysis clearly indicated that GLRaV-3 is present in all samples. SD samples were in most cases also co-infected with GVA and other vitiviruses notably GVE. Interestingly GVA from group II was not detected in the GLD control samples. The absence of GVA in one sample is a clear indication that GVA cannot be the only contributor to this disease along with GLRaV-3.

Table 1: List of reference sequences used in read mapping.

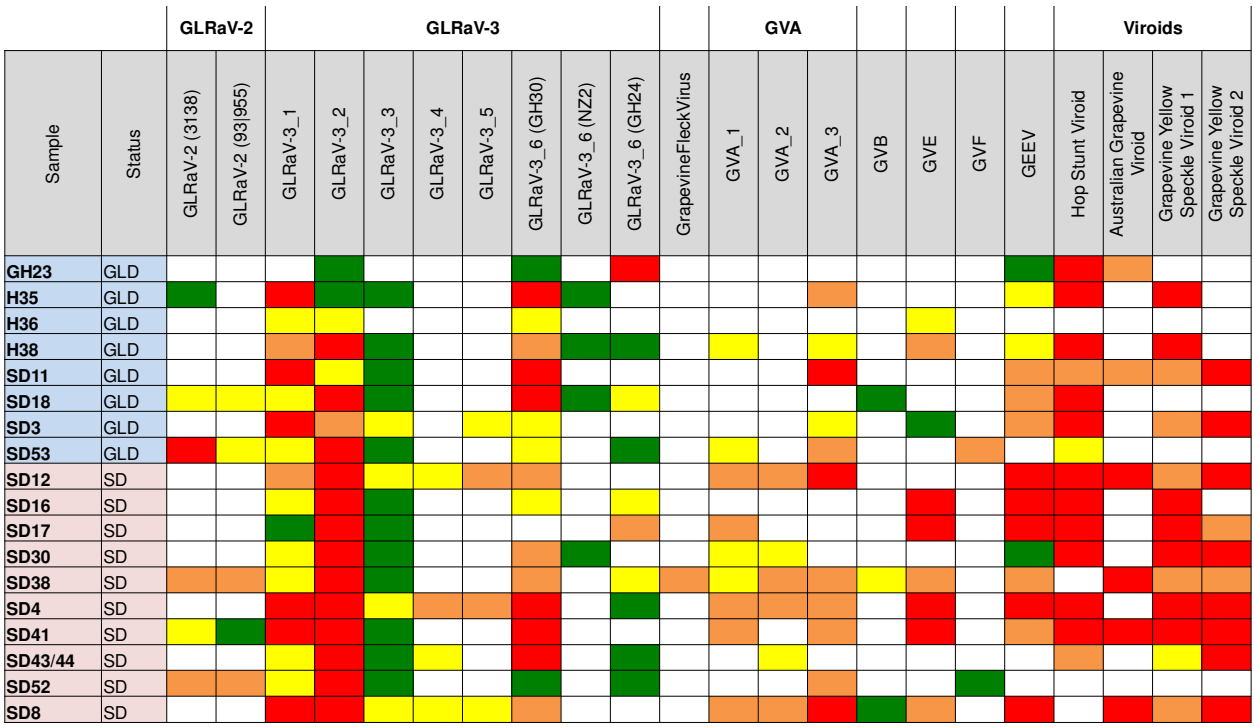
Common grapevine viruses	
Arabis Mosaic Virus	EU617326.1
Grapevine leafroll-associated virus 1	NC_016509.1
Grapevine leafroll-associated virus 2	JX559644.1
	AY881628.1
	DQ286725.2
Grapevine leafroll-associated virus 3	GQ352631.1
	EU259806.1
	GQ352633.1
	HQ401015.1
	HQ401016.1
	JQ655296.1
	KM058745
	JX220899.2
Grapevine leafroll-associated virus 4	NC_016416.1
Grapevine leafroll-associated virus 5	NC_016081.1
Grapevine leafroll-associated virus 6	NC_016417.1
Grapevine leafroll-associated virus 7	NC_016436.1
Grapevine leafroll-associated virus 9	AY297819.1
Grapevine Fanleaf Virus	KC900162.1
Grapevine Fleck Virus	NC_003347.1
Grapevine Red Blotch Associated Virus	NC_022002.1
Grapevine virus A	AF441234.2
	AF441235.2
	DQ855088.1
Grapevine virus B	NC_003602.1
Grapevine virus D	Y07764.1
Grapevine virus E	NC_011106.1
	GU903012.1
Grapevine virus F	NC_018458.1
Grapevine Rupestris associated Virus	AF026278.1
Tomato Spotted Wilt Virus	JF960237.1
Viroids	
Australian Grapevine Viroid	NC_003553.1
Grapevine Yellow Speckle Viroid 1	NC_001920.1
Grapevine Yellow Speckle Viroid 2	FJ597947.1
Citrus Exocortis Viroid	S67442.1
Hop Stunt Viroid	NC_001351.1

Minor grapevine viruses	
Grapevine Pinot gris virus	NC_015782.1
Grapevine Algerian Latent Virus	NC_011535.1
Grapevine Anatolian Ringspot Virus	NC_018383.1
Grapevine Angular Mosaic Virus	AY590305.1
Grapevine Asteroid Mosaic Associated Virus	AJ249357.2
Grapevine Berry Inner Necrosis Virus	NC_015220.1
Grapevine Bulgarian Latent Virus	NC_015492.1
Grapevine Chrome Mosaic Virus	NC_003622.1
Grapevine Deformation Virus	NC_017939.1
Grapevine Red Globe Virus	AF521977.1
Grapevine Rupestris Vein Feathering Virus	AY706994.1
Grapevine Syrah Virus 1	NC_012484.1
Grapevine Vein Clearing Virus	NC_015784.2
Rare grapevine viruses	
Alfalfa Mosaic Virus	NC_001495.1
Artichoke Italian Latent Virus	X87254.1
Blueberry latent virus	NC_014593.1
Blueberry Leaf Mottle Virus	U20621.1
Black raspberry cryptic virus	EU082132.1
Broad Bean Wilt Virus	EF528584.1
Carnation Mottle Virus	NC_001265.1
Cherry Leafroll Virus	NC_015414.1
Cucumber Mosaic Virus	FJ268744.1
Peach Rosette Mosaic Virus	AF016626.1
Petunia Asteroid Mosaic Virus	AY500881.1
Raspberry Bushy Dwarf Virus	NC_003739.1
Raspberry Ringspot Virus	NC_005266.1
Sowbane Mosaic Virus	HM163159.1
Strawberry Latent Ringspot Virus	NC_006964.1
Tobacco mosaic virus	NC_001367.1
Tobacco necrosis virus A	AY546104.1
Tobacco necrosis virus D	NC_003487.1
Tobacco Ringspot Virus	U50869.1
Tomato Black Ring Virus	NC_004439.1
Tomato Mosaic Virus	AF155507.2
Tomato Ringspot Virus	NC_003840.1

Grapevine associated mycoviruses	
Grapevine associated chrysovirus	GU108588.1
	GU108589.1
	GU108596.1
	GU108597.1
Grapevine endophyte endorna virus	JX678977
Grapevine associated mycovirus	GU108587
	GU108600.1
	GU108601.1
Grapevine associated narnavirus-1	GU108590.1
	GU108586.1
Mycoviruses	
Aspergillus fumigatus chrysovirus	FN178512.1, FN178513.2, FN178514.1, FN178515.1
Botryotinia fuckeliana partitivirus 1	NC_010349.1, NC_010350.1, NC_010351.1
Chalara endornavirus	GQ494150
Cryphonectria nitschkei chrysovirus 1	GQ290651
Diaporthe ambigua virus 1	AF142094.1
Discula destructiva virus 1	AF316992.1, AF316993.1, AF316994.2, AF316995.1
Discula destructiva virus 2	AY033436.1, AY033437.1
Fragaria chiloensis cryptic virus	NC_009519.1
Fusarium oxysporum chrysovirus 1	EF152346.1
Verticillium chryso virus	HM004067.2, HM004068.2, HM004069.2, HM004070.2
Vicia cryptic virus	EU371896.1
Heracleum Latent Virus	X79270.1
Hungarian Chrome Mosaic Virus	X15346.1
Ophiostoma quercus partitivirus	AM111099.1
Penicillium chrysogenum virus	NC_007539.1, NC_007540.1, NC_007541.1, NC_007542.1
Raphanus sativus cryptic virus 3	FJ461349.1, FJ461350.1
Rosa multiflora cryptic virus	EU024675.1, EU024676.1, EU024677.1
Rosellinia necatrix megabirnavirus	AB512282.1, AB512283.1

Table 2: Results of the most prominent viruses in the read-mapping analysis.

This document is confidential and any unauthorised disclosure is prohibited.



		GLRaV-2		GLRaV-3									GVA							Viroids			
Sample	Status	GLRaV-2 (3138)	GLRaV-2 (93|955)	GLRaV-3_1	GLRaV-3_2	GLRaV-3_3	GLRaV-3_4	GLRaV-3_5	GLRaV-3_6 (GH30)	GLRaV-3_6 (NZ2)	GLRaV-3_6 (GH24)	GrapevineFleckVirus	GVA_1	GVA_2	GVA_3	GVB	GVE	GVF	GEEV	Hop Stunt Viroid	Australian Grapevine Viroid	Grapevine Yellow Speckle Viroid 1	Grapevine Yellow Speckle Viroid 2
GH23	GLD																						
H35	GLD																						
H36	GLD																						
H38	GLD																						
SD11	GLD																						
SD18	GLD																						
SD3	GLD																						
SD53	GLD																						
SD12	SD																						
SD16	SD																						
SD17	SD																						
SD30	SD																						
SD38	SD																						
SD4	SD																						
SD41	SD																						
SD43/44	SD																						
SD52	SD																						
SD8	SD																						

Probability based on percentage of reference sequence covered

High	Medium	Low	Very Low	Neg
≥80%	<80% - ≥30%	<30% - ≥10%	<10%	0

This document is confidential and any unauthorised disclosure is prohibited.

Complete the following table

Milestone	Target Date	Extension Date	Date Completed	Achievement
Objective 1: To identify viruses (incl. variants) associated with SD.				
1a. Experimental Design and sample identification	2011		Completed	
1b. Optimize dsRNA extraction	2011		Completed	
2 Run#1	2011		Completed	
5 Run#2	Dec-12		Completed	
7 Run#3		Aug-13	Completed	
9. Analyse combined data.		Dec-13	Completed	No Association with specific viruses could be confirmed
6. Write scientific articles		Apr-13	Ongoing	
Objective 2: To design accurate diagnostic assays for SD.				
3. Design and optimize diagnostic RT-PCR for identified viruses	Oct-12		Completed	Diagnostic RT-PCR assays were developed for GEEV
Objective 3: To determine the biological contributions of individual viruses to the etiology of SD.				
4. Survey SD samples for identified viruses		Dec-13	Completed	No Association with specific viruses could be confirmed
8. Survey SD vines (field samples) for identified viruses		Aug-13	Completed	No Association with specific viruses could be confirmed
Objective 4: To plan effective disease control strategies.				**Not Achieved**

Accumulated outputs

Conclusions

The previous (existing) hypothesis stated that most (but not all) group II GVA variants are associated with SD. However, several samples (SD16, 17, 41, 52) did not test positive for GVA group 2 variants but were positive for GVE and/or GVA group III. No SD negative sample have been identified that were positive for the GVA group II variants. From a viral perspective the results would indicate that there might be multiple viruses or virus combinations that could trigger SD symptom expression of which GVA group II might be one. It seems likely that GVA (similar to isolate PA3, group III) and GVE can in some cases also elicit symptom expression.

In this study the association of a specific virus profile with SD could not be established and no new viruses were discovered that were consistently associated with the SD samples. However, GLRaV-3 was always present and this observation could be the key for control of this disease since no vitiviruses were found in the absence of GLRaV-3.

It would seem from the NGS results that the viral composition of SD is complex with undetermined interactions between known viruses. There is also the possibility of other factors such as host genetics or additional pathogens (bacterial, fungal or even phytoplasma) that could influence symptom expression. The exploration of other factors fell beyond the scope of this project but should be considered.

The most logical control for SD still remains the planting of clean material especially clean from GLRaV-3 and the rogueing of symptomatic plants.

This document is confidential and any unauthorised disclosure is prohibited.

Technology development, products and patents

All diagnostic primer sequences for RT-PCR assays are available. Since no association of a novel or known virus could be made conclusively, no technology transfer would be required.

Suggestions for technology transfer

Currently the planting of healthy material and removal of infected vines is the best control strategy and this practice should be promoted.

Human resources development/training

Student level (BSc, MSc, PhD, Post doc)	Cost to Project
1. BSc Hons (N Smyth 2012)	0
2. BSc Hons (J Rossouw 2013)	0
3. MSc (Y Espach 2011-12)	0
4. Postdoc (HJ Maree 2011-13)	0

Publications (popular, press releases, semi-scientific, scientific)

Espach, Y., Maree, H.J. and Burger, J.T. 2012. The complete genome of a novel endornavirus assembled from next-generation sequence data. Journal of Virology. 86(23), 13142.

Presentations/papers delivered

Conferences - International

1. 17th meeting of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), Davis, USA. 8-14 October 2012.
    - Maree, H.J., Espach, Y., Rees, D.J.G. and Burger, J.T. 2012. A study of Shiraz disease etiology using next-generation sequencing technology. p100. (Paper)
    - Espach, Y., Maree, H.J. and Burger, J.T. 2012. The use of next-generation sequencing to identify novel mycoviruses in single grapevine plants. p114. (Poster 23)
2. 22nd International Conference on Virus and Other Graft Transmissible Diseases of Fruit Crops (ICVF), Rome, Italy. 3-8 June 2012. [Petria 22 (3), p123-458 (2012)]
    - Maree, H.J., Nel, Y., Visser, M., Coetzee, B., Manicom, B., Burger, J.T. and Rees, D.J.G. 2012. The study of plant virus disease etiology using next-generation sequencing technologies. p283. (Paper)
3. Africa Virology Conference, Cape Town, South Africa. 29 November – 2 December 2011.
    - Maree, H.J., Coetzee, B., Nel, Y., Burger, J.T. and Rees, D.J.G. 2011. Unraveling the complexity of grapevine viral diseases using next-generation sequencing. (Poster 41)
    - Coetzee, B., Maree, H.J., Nel, Y., Rees, D.J.G. and Burger, J.T. 2011. The use of Next-Generation sequencing in metagenomic studies of plant viruses. (Paper)
4. Agricultural Biotechnology International Conference (ABIC), Sandton, South Africa. 6-9 September 2011.
    - Maree, H.J., Coetzee, B., Nel, Y., Burger, J.T. and Rees, D.J.G. 2011. Unraveling the complexity of grapevine viral diseases using next-generation sequencing. (Poster 5)

Conferences - National

1. 34th Conference of the South African Society for Enology and Viticulture, Franschhoek, South Africa. 14-16 November 2012.
    - Maree, H.J., Espach, Y., Smyth, N., Rees, D.J.G., and Burger, J.T. 2012. A study of Shiraz disease etiology using next-generation sequencing technology. (Paper)
    - Espach, Y., Maree, H.J., and Burger, J.T. 2012. The use of next-generation sequencing to identify novel mycoviruses in single grapevine plants. (Paper)

This document is confidential and any unauthorised disclosure is prohibited.

Workshops

1. Winetech virus workshop XI, Presentation: Study of grapevine viral diseases using next-generation sequencing technologies: A study of Shiraz and Leafroll disease (GenUS 10/1 and GenUS 11/3), Stellenbosch, South Africa. 29 May 2012.
2. Winetech virus workshop X, Presentation: A metagenomic sequencing approach to elucidate Shiraz disease aetiology (GenUS 10/1)

Total cost summary of the project

TOTAL COST IN REAL TERMS	COST		CFPA	DFTS	Deciduous	SATI	Winetech	THRIP	OTHER	TOTAL
YEAR 1							175439	87720		263159
YEAR 2							275000	137500		412500
YEAR 3							302500	151250		453750
YEAR 4										
YEAR 5										
TOTAL										**1129409**

This document is confidential and any unauthorised disclosure is prohibited.