Wild fish responses to wastewater treatment plant upgrades in the Grand River, Ontario

by

Kirsten Ellen Nikel

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Biology (Water)

Waterloo, Ontario, Canada, 2020

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Municipal wastewater treatment plant (WWTP) effluent is a major point source of contaminants (nutrients, pharmaceuticals, estrogens, etc.) which can harm aquatic life. Many studies have investigated the effects of WWTP effluent on male rainbow darter (*Etheostoma caeruleum*) collected downstream of two WWTPs in the Grand River, Ontario. These studies reported disruption at multiple levels of biological organization, including altered vitellogenin gene expression, lower levels of *in vitro* steroid production, and high frequency of intersex. The Region of Waterloo has invested in major upgrades at both WWTPs to improve effluent quality by increasing aeration and nitrification with the goal of reducing ammonia concentrations in the effluent. The Kitchener WWTP was initially upgraded in early 2013 with additional aeration/nitrification and extended solids retention time. After these upgrades, stable nitrogen isotope signatures in muscle tissue and *in vitro* steroid production of 11-ketotestosterone and testosterone in fish collected downstream of the outfall shifted to resemble upstream reference conditions, and there was a significant reduction in intersex incidence and severity. Upgrades to the Waterloo WWTP in 2017 and 2018 created a unique opportunity to investigate whether responses in rainbow darter previously associated with effluent exposure will resemble upstream reference levels following upgrades at a second WWTP. The biological endpoints in rainbow darter downstream of the Waterloo WWTP were not as severe as they were in fish downstream of the Kitchener WWTP but there was still potential for recovery. This thesis aimed to compare stable isotope signatures, *in vitro* steroid production, and intersex in rainbow darter caught upstream and downstream of the Waterloo WWTP before and after it was upgraded, to explore any changes that might occur following the upgrades. After the Waterloo WWTP upgrades were completed, there was a similar but less pronounced recovery and these endpoints were no longer statistically different from the upstream reference sites. However, it was often difficult to attribute these biological effects on fish directly to the changes in the WWTP effluent due to the variability in the endpoints among study sites, including the upstream reference sites. This unique long-term study is also valuable in explaining confounding effects of annual variations in water temperature and flow, as well as inputs from other sources which can mask or exacerbate the effects of the WWTP effluent. Overall, major capital investments in WWTP upgrades targeted at improving effluent quality have also corresponded with the reduction of adverse responses in fish.

Acknowledgements

My time as a graduate student at the University of Waterloo was both challenging and rewarding. Thank you to Dr. Mark Servos for being a wonderful supervisor, providing so many opportunities for personal and professional development, and always making science fun.

Thank you to Pam Law and Dominika Celmer-Repin at the Region of Waterloo who provided us with effluent datasets and gave us insight into the changes in effluent quality over the years. Your collaboration with us was essential for the success of this project and our ability to interpret our results.

Thank you to my committee members Dr. Gerald Tetreault and Dr. Paul Craig, who always asked interesting questions and provided great feedback.

I am grateful to have been part of a wonderful group of people within the Servos Lab. My fellow lab mates were an amazing support system and made the hard days better and the good days even more enjoyable. Thank you as well to Leslie Bragg and Hadi Dhieybi, who patiently answered my millions of questions, organized field work, and supported me in the lab by teaching me methods and troubleshooting with me. Thank you to all the co-op students, lab volunteers, and field volunteers who helped tremendously with gathering data and processing samples.

I would not have been able to do any of this work without the foundational studies done by Gerald Tetreault, Meghan Fuzzen, Keegan Hicks, and Patricija Marjan. Thank you for your guidance as I continued the research that you all started.

Finally, thank you to my family and partner for your endless support and encouragement.

Table of Contents

List of Figures

List of Tables

Chapter 1
General Introduction

Municipal wastewater treatment plant (WWTP) effluent is a major concern for the health of aquatic ecosystems, as it contains many contaminants that have been shown to cause a wide variety of adverse outcomes in aquatic species living downstream (Chambers et al. 1997; Vajda et al. 2008; Petrie et al. 2015). In the Grand River, the effects of wastewater effluent from two large treatment plants in Kitchener and Waterloo have been well studied. Wastewater exposure, and particularly exposure to natural and synthetic estrogens, has been found to be associated with reproductive system disruptions in a small bodied fish, rainbow darter (*Etheostoma caeruleum*), at multiple levels of biological organization. Reported effects include changes in stable isotope signatures (Loomer et al. 2015; Hicks et al. 2017b), gene expression (Bahamonde et al. 2014, 2015b), sex steroid production (Fuzzen et al. 2015, 2016), and the presence of intersex in male rainbow darter (Tetreault et al. 2011; Tanna et al. 2013; Bahamonde et al. 2015a). The Region of Waterloo has recently upgraded both the Kitchener and Waterloo WWTPs in order to improve effluent quality (i.e., reduce ammonia concentrations). The Kitchener WWTP's improved aeration and nitrification processes were completed between August 2012 and January 2013 (Region of Waterloo 2018) and these processes improved the effluent quality by lowering ammonia concentrations in the effluent and had a co-benefit of reducing effluent estrogenicity (Srikanthan 2019). Ongoing upgrades have continued, but after 2013 many adverse effects in rainbow darter in the receiving waters were reduced, including improved *in vitro* steroid production (Marjan et al. 2018) and lower incidence and severity of intersex (Hicks et al. 2017a). Though there was recovery in some endpoints after the Kitchener WWTP was upgraded, there are still gaps in our knowledge of how fish respond to various treatment plant upgrades over time. The upstream Waterloo WWTP also underwent major upgrades; however, there was a period of poor quality effluent during the ongoing construction (Region of Waterloo 2018). From 2009 to 2014 there was an increase in effluent ammonia due to construction delays. By 2015, return activated sludge (RAS) aeration was placed back online and the effluent quality began to improve (Pam Law, Region of Waterloo, personal communication). The final stages of the upgrades were completed after one aeration tank was finished in March 2017 and a second aeration tank came on line in March 2018 which promoted full year-round nitrification of the effluent. Although the biological effects (i.e., *in vitro* sex steroid production and intersex) in rainbow darter downstream of the Waterloo WWTP prior to the upgrades were not as severe as those associated with the Kitchener WWTP, improvements in effluent quality are expected to reduce adverse biological responses detected downstream of Waterloo. It is

important to determine how these major capital investments (more than $450M for the Kitchener and Waterloo WWTPs combined) in WWTP upgrades have affected the health of aquatic organisms in the environment in order to support future improvements in water management policy and practice. As studies on rainbow darter have been conducted for more than a decade in the Grand River (prior to the Waterloo WWTP upgrades) there was a unique opportunity to follow how major wastewater investments affect the biological responses in this sentinel species. This thesis focuses on the biological changes that occurred in rainbow darter in association with the Waterloo WWTP outfall during the period of the recent upgrades.

1.1 Wastewater effluent and emerging contaminants

A wide diversity of chemicals used in households and industry are collected by wastewater systems. WWTPs are important for the removal of these contaminants before effluent is discharged into the environment. In general, this process includes several steps (Chambers et al. 1997; Environment Canada 2001). Primary treatment involves physical processes such as mechanical bar screens to remove large debris. Solids are then settled by gravity in primary clarifiers. Next, the water passes through secondary (biological) treatment, where bacteria are grown/added to digest organic materials (Wang and Wang 2016). Biological solids settle by gravity within secondary clarifiers, and this sludge is pumped to anaerobic digesters for treatment along with the sludge from the primary clarifiers (Environment Canada 2001). Additional tertiary treatment may then be used to target and remove specific suspended or dissolved substances such as metals, organic chemicals, or nutrients still remaining in the water after secondary treatment. This can involve extra physical, chemical, or biological treatment. Finally, before release into the environment, the effluent can be disinfected by chlorination-dechlorination, UV light, or ozone.

These treatment processes are often not sufficient to remove all substances and contaminants of concern from municipal wastewater, and the effluent can continue to alter the environment in various ways (Wang and Wang 2016). Treated wastewater effluent contains nutrients such as nitrogen-based compounds and phosphorous that can lead to eutrophication (e.g., excessive algal growth) and reduce dissolved oxygen affecting the performance and survival of aquatic life (Chambers et al. 1997; Environment Canada 2001). Effluent can also increase the temperature of the receiving waters which can alter metabolism in aquatic organisms (Kaushal et al. 2010; Odjadjare and Okoh 2010; Mehdi et al. 2018). In addition, wastewater effluent often contains contaminants such as ammonia and metals that are

known to cause acute or chronic toxic effects in aquatic life (Chambers et al. 1997; Odjadjare and Okoh 2010). There has also been a growing focus on the effects of chemicals of emerging concern (CECs) (Daughton and Ternes 1999; Luo et al. 2014). CECs include a diversity of small, organic, compounds that WWTPs are not specifically designed to remove, and therefore they often occur at trace levels (ng/L to μg/L) in effluents (Luo et al. 2014). Despite their low concentrations, many of these CECs are continuously introduced into the environment and because of their potential to cause effects on growth, reproduction and development at very low doses they represent a risk to fish and other aquatic life (Daughton and Ternes 1999; Söffker and Tyler 2012; Matthiessen et al. 2018).

Pharmaceuticals and personal care products (PPCPs) are one class of CECs that have been found ubiquitously in municipal wastewater effluents and have the potential to exert subtle effects on the environment (Daughton and Ternes 1999; Luo et al. 2014). Originating from human, industrial, and agricultural use, these pharmaceuticals include chemicals such as antibiotics, beta blockers, antidepressants, and hormones (Corcoran et al. 2010; Wang and Wang 2016). These compounds are specifically designed to be biologically active, and since a high proportion of drug targets are evolutionarily conserved between humans and aquatic vertebrates, these compounds have the potential to affect biological function through a variety of mechanisms (Daughton and Ternes 1999; Gunnarsson et al. 2008). Natural (e.g., estrone (E1), 17β-estradiol (E2), and estriol (E3)) and synthetic (e.g., 17α-ethynylestradiol (EE2)) estrogens and other endocrine disrupting chemicals (EDCs) are of particular concern because they can be released via wastewater effluents into aquatic environments and can alter growth and reproduction in aquatic organisms such as fish at very low (< 10 ng/L) concentrations (Parrott and Blunt 2005; Jobling et al. 2006; Kidd et al. 2007).

The effects of treated wastewater effluent on aquatic species such as fish are still not fully understood (Mills and Chichester 2005; Corcoran et al. 2010; Söffker and Tyler 2012). While there is a great deal of evidence relating exposure to a single contaminant in the lab to adverse effects on fish health, there are challenges in linking these changes to wild populations (Mills and Chichester 2005). Wastewater effluent has been shown to affect fish at multiple levels of biological organization, including changes in gene expression such as the induction of vitellogenin gene expression in males (Liney et al. 2006; Vajda et al. 2008), changes in sex steroid production (Folmar et al. 1996; Hecker et al. 2002; Weber et al. 2019), the presence of intersex (Jobling et al. 2002; Vajda et al. 2008, 2011), a female-biased sex ratio (Vajda et al. 2008), changes to gonad weight (Jobling et al. 1998), reduced reproductive capacity (in severely intersex males) (Harris et al. 2011; Fuzzen et al. 2015), and changes in community structure (Tetreault et al. 2013; McCallum et al. 2019). These effects can depend on the species' sensitivity and life

history, reproductive strategies, and exposure to specific compounds or mixtures of compounds (Palace et al. 2009; Brown et al. 2014; Kidd et al. 2014). An organism's exposure to wastewater effluent can also be altered by river flow patterns and daily and seasonal variability in the effluent (Petrie et al. 2015).

1.2 Wastewater treatment in the Grand River

The Grand River watershed covers an area of 6,965 km^2 in southern Ontario and is the largest watershed flowing into the Canadian side of Lake Erie (Loomer and Cooke 2011). The watershed is important in many ways to its almost one million inhabitants, as it provides drinking water, is used recreationally for boating and fishing, and assimilates wastewater from 30 WWTPs and uncounted septic fields. The river is actively managed with dams and reservoirs to prevent flooding and to maintain summer low flow conditions for wastewater dilution (Loomer and Cooke 2011). The watershed hosts a great deal of agricultural activity (~70% of total land use). In particular, the Conestogo River, which enters the Grand River just north of Waterloo, has some of the highest agricultural production, livestock density, and tile drainage systems in the watershed. The central region of the Grand River, which includes the cities of Waterloo and Kitchener, has the highest urban use and population density. The central Grand receives inputs from nine WWTPs, including the largest two in the watershed located in Kitchener and Waterloo (Loomer and Cooke 2011). The Kitchener and Waterloo WWTPs have been the focus of many studies on key reproductive endpoints over the past several years as they have both undergone various upgrades (summarized in Table 1.1).

Table 1.1 Overview of treatment processes and upgrade history at the two largest WWTPs on the Grand River, located in Waterloo and Kitchener (updated from Hicks et al. 2017b plus additional sources listed below).

	Waterloo WWTP			**Kitchener WWTP**	
Year	2007–2009 (pre-upgrade)	2009–2016 (during upgrades)	2017–present (post-upgrade)	2007–2012 (pre-upgrade)	2013–present (post-upgrade*)
Population served	138,464 (2016) [2]		153,902 (2019) [2]	227,761 (2011) [2]	256,513 (2019) [2]
Measured Flow (m^3/d)	39,750 (2016) [2]		41,805 (2018) (2019 data not available) [2]	70,443 (2011) [2]	67,902 (2018) (2019 data not available) [2]
Rated treatment capacity (m^3/d)	54,600–57,500 [2]		57,500 [2]	122,700 [2]	122,700 [2]
Treatment Process	Conventional activated sludge [1]			Conventional activated sludge[1]	
Primary Treatment	Screening, grit removal, ferric sulphate for phosphorous removal, primary clarification. Ferric chloride added upstream of the primary clarifiers and/or aeration tanks for phosphorus removal[1]			Screening, grit removal, primary clarification, aeration. Ferric chloride upstream of the primary clarifiers and ferrous chloride downstream of the primary clarifiers. [1]	
Secondary Treatment	Aeration, secondary clarifier[1]	Partial nitrification from 2009 to 2014[1]	Increased aeration, full year-round nitrification[1]	Aeration, secondary clarifier[1]	Aeration tanks to enhance ammonia removal[1]
Tertiary Treatment	NA[1]		NA[1]	NA[1]	August 2017: Tertiary Filters Commissioned (Plant 1 not being filtered)[3]
Disinfection	Sodium hypochlorite;		UV light[1]	Sodium hypochlorite; sodium	UV light[1]

	sodium bisulphite de-chlorination [1]		bisulphite de-chlorination [1]	
Biosolids		Anaerobic digestion, dewatered, used in agriculture or mine remediation or sent to landfill[1]	Lagoon storage[1]	Anaerobic digestion, dewatered, used in agriculture or mine remediation or sent to landfill[1]

*A major new treatment plant is anticipated to come online in 2019–2020 (plant 3) and the original small plant (plant 1) will be decommissioned. The upgraded plant 2 will continue to operate. (Pam Law, Region of Waterloo, personal communication)
[1] (Region of Waterloo 2018)
[2] (Region of Waterloo 2019)
[3] (Pam Law, Region of Waterloo, personal communication)

Both the Kitchener and Waterloo WWTPs play an important role in nutrient loading and water quality in the Grand River. The two WWTPs contribute approximately 10% and 5% of the total river flow at their discharge points, respectively (Arlos et al. 2018). During the low flow summer months, point source loadings from the Kitchener and Waterloo WWTPs contribute about 70% of the phosphorus in the central Grand River. Conversely, during the spring months, point sources only contribute 3% of the phosphorus, and runoff from non-point (agricultural) sources is the primary driver of phosphorus loadings (Grand River Conservation Authority 2014). Total nitrate, total phosphorous, and chloride concentrations generally increase as the Grand River flows downstream through Waterloo and Kitchener (Loomer and Cooke 2011). Wastewater effluent must adhere to the Wastewater Systems Effluent Regulations, which sets limits for carbonaceous biological oxygen demand, total suspended solids, chlorine, unionized ammonia, as well as acute toxicity in effluents (Ministry of Justice 2012).

Before the Kitchener WWTP was upgraded, there was minimal nitrification, inefficient aeration and a short solids retention time (SRT) of < 2 days (Hicks et al. 2017a). After the Kitchener WWTP was upgraded, aeration was increased, SRT was improved to > 5 days (plant 2), and ammonia concentrations in the final effluent were reduced from 25 mg/L to 2–6 mg/L. In addition, pharmaceuticals such as ibuprofen and naproxen were reduced in the final effluent (Hicks et al. 2017a). Dissolved oxygen in the receiving environment also improved from as low as 1.0 mg/L during the summer before upgrades to generally > 6 mg/L in the summers after the upgrades (Hicks et al. 2017a). Although the treatment plant upgrades were primarily designed to meet effluent quality objectives (Region of Waterloo 2018), the improved nitrification processes also contributed to the reduction in estrogenicity of the effluent (Hicks et

al. 2017a). Subsequently, positive responses have been observed in wild fish caught downstream of the outfall of the Kitchener WWTP, the first of the two facilities to be upgraded (Hicks et al. 2017a; Marjan et al. 2018).

The Waterloo WWTP initially had partial nitrification and lower ammonia concentrations in its effluent compared to the Kitchener WWTP, but ammonia concentrations increased during its upgrade period because the aeration upgrades were not completed (due to a construction delay), and therefore the plant was not fully nitrifying from 2009 to 2014 (Hicks et al. 2017a). Concentrations of estrogens in the Waterloo WWTP effluent were also generally lower than Kitchener but more variable (Hicks et al. 2017a). Prior to its upgrades, the Waterloo WWTP was estimated to remove E1, E2, and EE2 at rates of 55%, 95%, and 69%, respectively (Arlos et al. 2018). As the aeration tank upgrades were completed at the Waterloo WWTP, full year-round nitrification was achieved and ammonia concentrations and estrogenicity in the final effluent were greatly reduced (Srikanthan 2019). A question of interest for wastewater managers was how wild fish would respond to these changes in effluent quality after the infrastructure upgrades.

1.3 Rainbow darter in the Grand River

The Waterloo and Kitchener WWTPs have been the focus of many studies over the past several years, and particular attention has been given to reproductive endpoints in a sentinel fish species, the rainbow darter (*Etheostoma caeruleum*) (Tetreault et al. 2011; Tanna et al. 2013; Bahamonde et al. 2015a, b; Fuzzen et al. 2015; Hicks et al. 2017a; Marjan et al. 2018). Rainbow darter are a small-bodied, benthic species of fish found in riffle habitats (Reeves 1907; Winn 1958), and are native to the Grand River watershed (Tetreault et al. 2011). A community survey found that rainbow darter were the most abundant species caught along most reaches of the central Grand River (Tetreault et al. 2013). These fish are sexually dimorphic, short-lived (five years), reach sexual maturity at one year, and spawn in the spring (asynchronous clutch spawners; Reeves 1907; Fuzzen et al. 2016; Hicks et al. 2017a). Rainbow darter have a small home range which has been confirmed through mark-recapture (Hicks and Servos 2017) and are therefore an ideal sentinel study species.

1.3.1 Timing of rainbow darter sampling

It is important to choose a suitable time for fish sampling to reduce variability in the measured endpoints in order to improve the study's ability to detect differences without increasing sample sizes (Galloway

and Munkittrick 2006). Rainbow darter spawn in early spring, from April to May and temperature is an important factor for both male and female gonad development, as spawning has not been observed below temperatures of 15°C (Reeves 1907; Winn 1958). Since temperature affects the exact timing of their spawn, it is challenging to sample fish at a consistent time point during their spawning period, year after year, especially when access to sites is often limited due to increased flows from snowmelt (Fuzzen et al. 2016). Rainbow darters may also potentially exhibit greater movement during this period in order to find suitable spawning habitat and food (Winn 1958; Hicks and Servos 2017). Consequently, more variability in many endpoints has been observed when fish are sampled in the spring compared to the fall (Fuzzen et al. 2016). Therefore, it is desirable to sample rainbow darter in the fall when there is limited in-stream movement and there has been substantial investment in gonadal development but the endpoints (e.g., *in vitro* sex steroid production) demonstrate less variability (Fuzzen et al. 2016; Marjan et al. 2018). Barrett and Munkittrick (2010) made a similar recommendation for early spring spawning fish that are difficult to sample 4–6 weeks pre-spawn as a part of their review of the Environmental Effects Monitoring (EEM) program data.

1.4 Exposure and responses to wastewater effluent in the Grand River

Exposure to wastewater in the Grand River has been shown to affect wild rainbow darter at all levels of biological organization. Various methods have been used to assess exposure of fish to WWTP effluent in the receiving environment, including changes in stable isotope signatures (Loomer et al. 2015; Hicks et al. 2017b). Key reproductive responses that have been measured in rainbow darter in association with wastewater effluent exposure are reductions in *in vitro* sex steroid production (Fuzzen et al. 2015, 2016) and high levels of intersex (Tetreault et al. 2011; Tanna et al. 2013; Bahamonde et al. 2015a).

1.4.1 Stable isotopes

Stable isotopes are generally used to discern how energy flows through food webs and to identify the source of nutrient inputs (Peterson and Fry 1987; Post 2002; Jardine et al. 2006). Isotopic composition of a sample is reported in δ notation, describing a parts per thousand difference from a standard. $\delta X = [(R_{sample} / R_{standard}) - 1] \times 10^3$ where X is the heavy isotope of carbon or nitrogen (^{13}C or ^{15}N, respectively) and R is the ratio of heavy to light isotopes in the sample or standard. Isotopic signatures are a reflection of the assimilation of nutrients in the organism over weeks or months (Angradi 1994). Stable isotope signatures have been often been used as indicators of sewage contamination in stream systems (Steffy and

Kilham 2004; Morrissey et al. 2013). Isotopic signatures have been shown to change dramatically in fish near wastewater effluent outfalls in the Grand River and can be linked to effluent quality (Loomer et al. 2015; Hicks et al. 2017b).

Carbon does not fractionate much as it is processed through the food web (only 0.4 ± 1.3‰ per trophic level; Post 2002). The carbon isotopic signature is therefore retained and can be useful for identifying the source of the carbon (Peterson and Fry 1987). Carbon isotopic signatures from aquatic sources can range from −40 to −20‰, whereas carbon from terrestrial sources (e.g., introduced via WWTP effluent) has a $\delta^{13}C$ of approximately −28‰ (France 1995a). However, due to overlap in these values as well as influence from various environmental conditions it is often difficult to form strong conclusions regarding carbon sources in aquatic systems (France 1995a). Rainbow darter caught downstream of both the Waterloo and Kitchener WWTPs have been shown to be enriched in $\delta^{13}C$ compared to those caught at upstream sites, indicating that fish at these sites are assimilating sewage-derived nutrients (Hicks et al. 2017b). $\delta^{13}C$ signatures in the Grand River have also been shown to increase gradually moving downstream and in response to dams (Loomer 2008; Hicks et al. 2017b). $\delta^{13}C$ signatures also vary annually depending on flow conditions in the river, where years with lower flows tend to have more enriched $\delta^{13}C$ signatures (Hicks et al. 2017b). Despite these other factors, the pattern of $\delta^{13}C$ enrichment remained similar both before and after the Kitchener WWTP upgrades, indicating that the level of wastewater treatment did not affect $\delta^{13}C$ in rainbow darters (Hicks et al. 2017b). Downstream of the Waterloo WWTP, $\delta^{13}C$ was also enriched in rainbow darter compared to upstream reference sites (Hicks et al. 2017b), and this pattern is expected to remain after the WWTP upgrades.

Nitrogen generally has a mean fractionation of 3.4 ± 1‰ per trophic level and can therefore be used to determine an organism's trophic position (Post 2002; Jardine et al. 2006). This enrichment is primarily due to organisms' excretion of isotopically light nitrogen (Peterson and Fry 1987; Ulseth and Hershey 2005). There is a difference in isotopic signatures between organic matter originating from aquatic sources compared to terrestrial sources (Jardine et al. 2006; Brown et al. 2011). There is a positive relationship between the human population of a watershed and $\delta^{15}N$ signatures due to the high $\delta^{15}N$ content of human sewage (Cabana and Rasmussen 1996) and there is also a positive correlation between $\delta^{15}N$ and nitrogen loadings from synthetic fertilizers and livestock populations (Anderson and Cabana 2006). Wastewater inputs can alter isotopic signatures of primary producers, benthic invertebrates, and fish in aquatic environments (Cabana and Rasmussen 1996; Jardine et al. 2006; Brown et al. 2011). Lack of nitrification and denitrification (i.e., raw sewage or primary treatment) results in ammonia being depleted in ^{15}N and therefore primary consumers (i.e., invertebrates) become depleted in ^{15}N which is then

reflected in their fish consumers (Anderson and Cabana 2006; Rasmussen and Trudeau 2010; Hicks et al. 2017b). With nitrification and denitrification (secondary treatment), more ^{15}N is released and the $\delta^{15}N$ signature in primary consumers and fish is enriched (Hicks et al. 2017b). In the Grand River, $\delta^{15}N$ signatures have been found to be correlated with total ammonia released, where lower volumes of ammonia correspond to more enriched $\delta^{15}N$ signatures (Hicks et al. 2017b). Prior to the upgrades at the Kitchener WWTP, $\delta^{15}N$ signatures in rainbow darter were found to be depleted at downstream sites compared to upstream sites (Loomer et al. 2015; Hicks et al. 2017b). After the Kitchener WWTP was upgraded, $\delta^{15}N$ signatures increased downstream of the outfall in rainbow darter and were indistinguishable from isotopic signatures upstream (Hicks et al. 2017b). A similar pattern was seen in macrophytes, epilithon, and seston (Cejudo et al. 2018) as well as primary consumers (Hicks et al. 2017b) before and after the Kitchener WWTP upgrades. Downstream of the Waterloo WWTP there was a slight enrichment in $\delta^{15}N$ in rainbow darter at downstream sites during years of better quality effluent (2007–2010) but depletion during construction years when effluent quality was poor (2011–2014; Loomer et al. 2015; Hicks et al. 2017b). $\delta^{15}N$ signatures in rainbow darter after the 2017–2018 upgrade to the Waterloo WWTP are not yet known but isotopic enrichment is expected to occur.

1.4.2 *In vitro* steroid production

Sex steroids (estrogens and androgens) play an important physiological role in all species, as they are important for sexual differentiation and sexual development (Tyler et al. 1998; Devlin and Nagahama 2002; Söffker and Tyler 2012). In fish, the main androgens are 11-ketotestosterone (11KT) and testosterone (T) (Tyler et al. 1998). Lower levels of sex steroid production have been associated with estrogenic WWTP effluent exposure in many studies (Folmar et al. 1996 - serum; Tyler et al. 1998 - serum; Hecker et al. 2002 - circulating; Weber et al. 2019 - steroids measured in homogenized gonad). Prior to the Kitchener WWTP upgrades, rainbow darter caught downstream of the outfall had significantly lower stimulated *in vitro* 11KT and T production compared to fish caught from upstream reference sites (Fuzzen et al. 2015, 2016).

Within two years after the Kitchener WWTP upgrades, 11KT production in males caught downstream was no longer reduced compared to males caught from reference sites (Marjan et al. 2018). *In vitro* production of T was more variable, but showed similar patterns to 11KT, increasing downstream of the Kitchener WWTP after the upgrades. In 2013 and 2014, a reduction in 11KT was observed at a reference site located 4.7 km upstream of the Waterloo WWTP outfall. The reason for this was unclear, so additional sites near the Conestogo River confluence were added in 2015; however, the effect was not

seen that year (Marjan et al. 2018). This suggests that there may be additional contaminants that can enter the system, or that steroid production is dependent on some factors that are still not understood.

Compared to the Kitchener WWTP, effluent quality from the Waterloo WWTP was more variable over time (from 2009 to 2014) and fish caught downstream showed more variable production of 11KT and T. Recovery in 11KT production after the Kitchener WWTP upgrade was gradual (Marjan et al. 2018), so it is expected that a recovery downstream of the Waterloo plant may not be observed immediately. There was a tendency for reduced steroid production downstream of the Waterloo WWTP compared to the two farthest upstream reference sites, but overall the impact of wastewater effluent was less evident at Waterloo compared to the Kitchener outfall, possibly due to differences in the effluent quality and exposure (Marjan et al. 2018). Lower *in vitro* steroid production in some years at the immediate upstream reference site may make detection of changes below the outfall difficult to detect and interpret as steroid production may be already partially suppressed due to unknown confounding factors.

1.4.3 Intersex incidence and severity

Intersex, or the simultaneous presence of both male and female tissue in the gonads in a fish of a gonochoristic (fixed-sex) species, is often associated with exposure to EDCs (Jobling et al. 1998; Tyler and Jobling 2008; Bahamonde et al. 2013). Intersex incidence and severity have been observed in rainbow darter in the Grand River downstream of both the Kitchener and Waterloo WWTPs (Tetreault et al. 2011; Tanna et al. 2013). In studies on the Grand River, male intersex severity was scored on a scale ranging from 0 (100% male tissue) to 7 (100% female tissue; Bahamonde et al. 2015a). Intersex incidence (%) was reported as the number of male fish with intersex divided by the total number of fish sampled (Fuzzen et al. 2016; Hicks et al. 2017a). Intersex is a concern because it has been found to be related to a reduction in reproductive capacity (Harris et al. 2011; Fuzzen et al. 2015). Severely intersex male rainbow darter (index 4–6) had significantly lower fertilization success compared to normal (index 0) or moderately feminized (index 1–3) males (Fuzzen et al. 2015).

Prior to the WWTP upgrades, 70–100% of rainbow darter caught downstream of the Kitchener WWTP exhibited intersex, with a mean score of 2–3 (Hicks et al. 2017a). In the first fall season post-upgrade, intersex incidence decreased to 29%, and by the third season it was 9–14%. Intersex severity also decreased such that by the third fall season the mean intersex score was less than one. Intersex severity was observed to recover faster compared to intersex incidence, but throughout all sampling periods, incidence was positively correlated with severity (Hicks et al. 2017a). Compared to the Kitchener WWTP, rainbow darter caught downstream of the Waterloo WWTP did not exhibit intersex as often

(ranging from 10% to 55%) or with as much severity (mean score of around one). This is likely explained by the lower volume and higher quality, although variable, effluent coming from the Waterloo WWTP throughout the study period (Hicks et al. 2017a).

Recovery of intersex after the Kitchener WWTP upgraded was likely linked to a reduction of estrogenicity in the wastewater effluent from Kitchener, as an indirect effect of the improved nitrification process (Hicks et al. 2017a). The upgrades to the Kitchener plant were initiated in mid 2012, but reduction in intersex was observed in the fall of 2013, after a full year of improved water quality. Since rainbow darter undergo recrudescence (period of gonadal growth) in the fall for the next spring spawning period, the time of year that fish are exposed to EDCs may influence the development of intersex (Hicks et al. 2017a). It is expected that intersex incidence and severity will be reduced downstream of the Waterloo WWTP after the treatment upgrades, and it might take a full year or more to observe responses in rainbow darter due to the timing of the WWTP upgrades in relation to the critical window of gonadal development.

1.5 Research Objectives

Emerging contaminants such as EDCs and PPCPs found within wastewater effluent have been identified as one of the top threats to freshwater biodiversity, and improving wastewater treatment is one method of alleviating the problem (Reid et al. 2018). Long-term ecological studies are extremely important for broadening our understanding of the effects that humans have on the environment, and for guiding policy and management decisions to mitigate these impacts (Hughes et al. 2017). Long-term studies, including this one, will become even more useful in the future as the environment is rapidly changing due to climate change and urbanization (Hughes et al. 2017). This study is unique because it is one of only a few studies conducted over a long period of time on both the ecological effects of wastewater effluent and how these effects change with improvements in effluent quality.

Major upgrades were completed at the Kitchener WWTP in early 2013 and the Waterloo WWTP in 2017–2018, providing a unique opportunity to assess how reproductive endpoints in rainbow darter responded to an improvement in water quality (Region of Waterloo 2018). Since 2013, recovery of many endpoints has been observed in rainbow darter downstream of the Kitchener WWTP (Hicks et al. 2017a; Marjan et al. 2018). Key responses included a dramatic recovery of the severity and incidence of intersex (Hicks et al. 2017a) and sex steroid production (Marjan et al. 2018) after the upgrades were complete. It is anticipated that upgrades to the Waterloo WWTP in 2017–2018 will lead to similar positive responses.

This research is especially valuable since it has the potential to reaffirm confidence in the success of the Kitchener WWTP upgrades, as well as to evaluate if the large capital investment in the upgrades of the Waterloo WWTP to improve effluent quality also corresponds to improved water quality in the Grand River and the health of aquatic life. The Waterloo WWTP processes a lower volume of wastewater (~62% of the volume treated by the Kitchener WWTP; Tetreault et al. 2013) and had historically higher quality effluent than the Kitchener WWTP (although there was reduced effluent quality during the period of construction). Additionally, the adverse effects on rainbow darter (e.g., sex steroid production, intersex) were not as severe as they were downstream from the Kitchener WWTP. The fish downstream of the Waterloo WWTP also have more opportunity to move out of the direct effluent plume since the effluent is discharged at the side of the river, whereas at the Kitchener WWTP the downstream environment is more dominated by the effluent due to a diffuser. For these reasons, the effects of the upgrades at Waterloo may be harder to detect than the effects of the upgrades at Kitchener. However, it is important to determine if the major infrastructure investments made at the Waterloo WWTP will lead to improvements in the responses in fish to support future management decisions.

This study aimed to further investigate the responses of rainbow darter to major WWTP infrastructure upgrades in the Grand River. The fall of 2018 was the first year after the Waterloo WWTP upgrades were completed, and therefore the fish were exposed to higher quality effluent for a full reproductive cycle, including their critical window for gonadal development. The objective of this research was to determine if there was a change in exposure (indicated by stable isotope signatures) or key biological endpoints (*in vitro* sex steroid production and intersex) downstream of the Waterloo WWTP after the major upgrades.

Chapter 2
Responses of rainbow darter to the Waterloo WWTP upgrades

2.1 Introduction

Treated wastewater effluent is a complex stressor for aquatic ecosystems characterized by ammonia, excess nutrients, biological oxygen demand, suspended solids, metals, pharmaceuticals, and natural and synthetic hormones which can cause both acute and chronic effects (Luo et al. 2014; Petrie et al. 2015). While some biologically active compounds are readily degraded during the wastewater treatment process, others are persistent and are less easily removed (Rojas et al. 2013). Endocrine disrupting compounds (EDCs) are of particular concern as their removal rates from wastewater effluent vary widely and they are frequently detected in the environment in the low ng/L range (Ternes et al. 1999; Servos et al. 2005).

Reproduction in wild fish collected downstream of wastewater outfalls has been disrupted at all levels of biological organization, and many of these disruptions have been linked to estrogens and EDCs. Wild fishes' specific responses to estrogens can depend on their life history (Palace et al. 2009), life stage during their exposure (van Aerle et al. 2002), as well as site-specific environmental conditions that modulate their exposure such as flow (Barber et al. 2012). Wild fish exposure to wastewater effluent has been shown to be associated with changes in gene expression such as the induction of vitellogenin gene expression in males (Liney et al. 2006; Vajda et al. 2008), alterations in endogenous sex steroid production (Folmar et al. 1996; Hecker et al. 2002; Weber et al. 2019), the induction of intersex (Jobling et al. 2002; Vajda et al. 2008, 2011), changes in gonad weight (Jobling et al. 1998), reduction in reproductive capacity (in severely intersex males) (Harris et al. 2011; Fuzzen et al. 2015), and changes in community structure (Tetreault et al. 2013; McCallum et al. 2019). Effects at higher levels of biological organization (e.g., population and community changes) are often harder to link to specific components of effluents. Laboratory studies have shown changes in growth and condition and a reduction in fertilization success in response to EE2 exposure (Parrott and Blunt 2005). As well, the collapse of a fish population was reported after an entire lake was dosed with EE2 (Kidd et al. 2007) and there were also indirect effects on other organisms in that food web (Kidd et al. 2014). Taken together, these effects show that there is the potential for EDCs in WWTP effluent to exert harm on fish populations either alone or in combination with other constituents of the effluent.

Many reproductive effects have been reported in wild rainbow darter (*Etheostoma caeruleum*) living downstream of WWTPs in the Grand River, Ontario. The Grand River watershed is the largest watershed draining into the Canadian side of Lake Erie and covers an area of 6,965 km^2 (Loomer and

Cooke 2011). The watershed is home to nearly one million people, assimilates water from 30 WWTPs, is highly agricultural, and its flow is actively managed by dams and reservoirs (Loomer and Cooke 2011; Grand River Conservation Authority 2014). Two of the largest WWTPs in the watershed, located in Kitchener and Waterloo, have been the focus of many studies investigating the effects of WWTP effluent on wild fish.

In rainbow darter living downstream of the Waterloo and Kitchener WWTPs, exposure to excess nutrients in wastewater effluent has been shown through changes in stable isotope signatures (Loomer et al. 2015; Hicks et al. 2017b) and effluent estrogenicity has been associated with increased levels of vitellogenin gene expression in males (Bahamonde et al. 2014, 2015b). Reproductive impairments have also been observed, including reductions in *in vitro* sex steroid production of 11-ketotestosterone and testosterone (Fuzzen et al. 2015, 2016) and high incidence of intersex (Tetreault et al. 2011; Tanna et al. 2013; Bahamonde et al. 2015a).

Both the Kitchener and Waterloo WWTPs are secondary-level conventional activated sludge plants and have recently undergone major process upgrades aimed at improving effluent quality by increasing aeration/nitrification (Region of Waterloo 2018). The Kitchener WWTP serves a population of ~250,000 and treats an average volume of 67,902 m^3 per day (Region of Waterloo 2019). The main upgrades to the Kitchener WWTP were completed in 2013 and included new aeration tanks to enhance ammonia removal. There was also a co-benefit of reducing estrogenicity in the effluent (Hicks et al. 2017a; Arlos et al. 2018; Srikanthan 2019). These upgrades were associated with a reduction of rainbow darter exposure to ammonia (assessed via stable isotopes; Hicks et al. 2017b) and recovery of *in vitro* steroid production (Marjan et al. 2018) and intersex (Hicks et al. 2017a). The Waterloo WWTP serves a population of ~150,000 and treats a volume of 41,805 m^3 per day (Region of Waterloo 2019). Upgrades to the Waterloo WWTP began in 2009 but there were construction delays and thus there was a period of time from 2009 to 2014 during which there was only partial nitrification of the effluent (Table 2.1). Once RAS re-aeration came back online in 2015, ammonia levels began to decrease, and once the aeration tanks were upgraded in 2017 and 2018, the plant achieved full year-round nitrification (Region of Waterloo 2018). The objective of the current study is to examine the effects of the Waterloo WWTP upgrades on stable isotopes, *in vitro* sex steroid production, and intersex in wild male rainbow darter.

Table 2.1 Timeline of the Waterloo WWTP upgrades (Srikanthan 2019; Pam Law, Region of Waterloo, personal communication).

Upgrade Description	Commissioning Date
Interim Dewatering (note: centrate sent to the Raw Sewage PS, aeration upgrades not completed, so temporary increase in effluent ammonia)	2009–2014
UV disinfection	2012
RAS re-aeration online	2015
Aeration tank 1 upgrades	March 2017
Aeration tank 2 upgrades	March 2018

2.2 Methods

2.2.1 Study species

Rainbow darter (*Etheostoma caeruleum*) are a small-bodied fish native to and abundant in the Grand River. These fish are sexually dimorphic, short-lived (five years), reach sexual maturity at one year, and spawn in the spring (Reeves 1907; Fuzzen et al. 2016; Hicks et al. 2017a). They have a small home range which has been confirmed through mark-recapture (Hicks & Servos 2017) and are therefore an ideal sentinel study species to examine the spatial effects of wastewater effluent.

2.2.2 Site description

To assess responses in rainbow darter over the course of WWTP upgrades, five sites were selected along an urban gradient throughout the central Grand River both upstream and downstream of the Waterloo WWTP. Three reference sites were chosen in upstream rural and urban environments (Figure 2.1). The farthest upstream site (REF1; 43°37'51.07"N, 80°26'36.68"W) was located within a rural area 31.8 km upstream of the WWTP outfall and 18 km downstream of Shand Dam, a large bottom draw dam operated by the Grand River Conservation Authority for flood protection and flow augmentation (Grand River Conservation Authority 2014). The second reference site (REF2; 43°35'10.34"N, 80°28'47.77"W) was chosen 21.5 km upstream of the outfall and 8.9 km upstream of the Conestogo River's confluence with the Grand. The Conestogo River is a large agricultural input into the Grand River and contains a reservoir approximately 38 km upstream of its confluence with the Grand River (Loomer and Cooke 2011). The third reference site (REF3; 43°30'16.50"N, 80°28'34.52"W) was located 7.8 km downstream of the

Conestogo River and 4.8 km upstream of the outfall, within the urbanized area of the City of Waterloo. Additional sites near the Conestogo River confluence (12.5 km upstream of the wastewater outfall) were sampled in 2015 and 2018 in order to assess any confounding effects that agricultural activity in the Conestogo River might be having on fish living at REF3 (Marjan et al. 2018; Appendix A2). The first downstream site (DSW; 43°28'24.44"N, 80°28'23.32"W) was located 1.0 km downstream of the WWTP outfall. The second downstream site (INT1; 43°26'38.14"N, 80°24'2.43"W) was located 12.0 km downstream of the outfall. All sites consisted of riffle/run habitat.

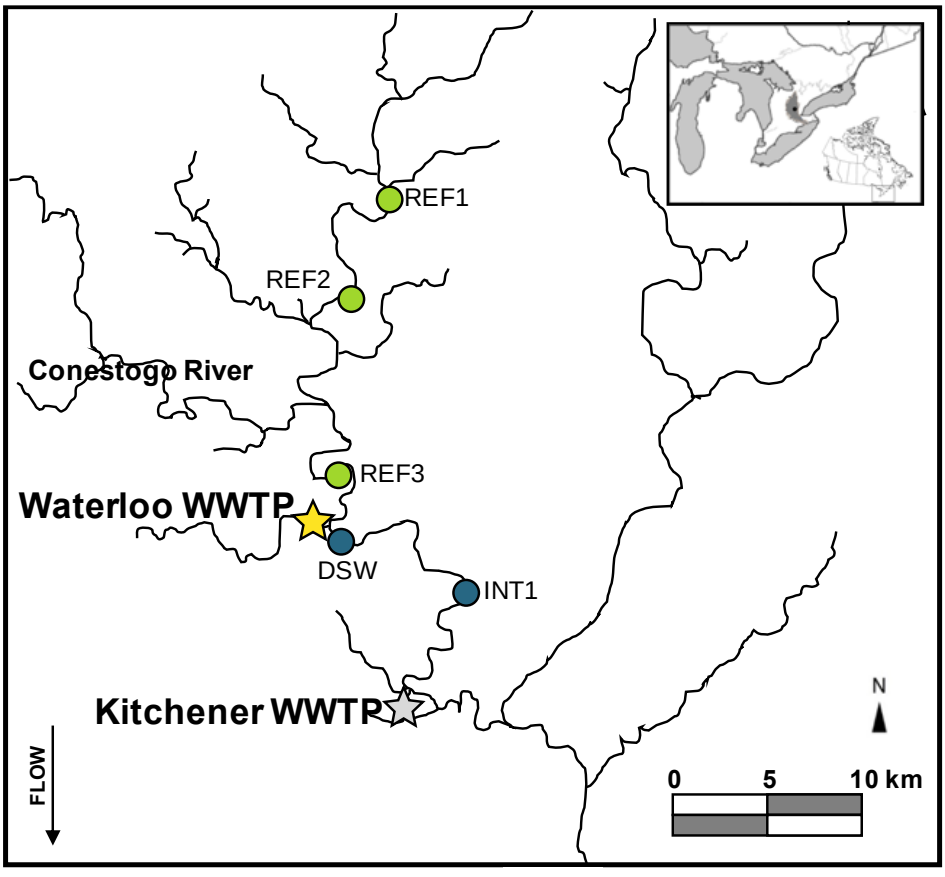

Figure 2.1 Map of sampling locations along the Grand River. Three reference sites were located upstream of the Waterloo WWTP and two sites were sampled downstream. The Kitchener WWTP is indicated by the grey star for context and is located approximately 21.2 km downstream of the Waterloo WWTP.

2.2.3 Fish collection and processing

Rainbow darter were collected in late October of each year (2007 and 2010–2019), which was suggested by Barrett and Munkittrick (2010) and Fuzzen et al. (2016) to be an appropriate time in the darter's reproductive cycle for sampling as it is difficult to obtain consistent samples 4–6 weeks pre-spawn. In a supplementary study, this time period was confirmed to be suitable for measuring *in vitro* steroid production in rainbow darter since steroid production was consistent throughout the months of October to December (Marjan et al. 2018), and isotope signatures in rainbow darter were also shown to be more sensitive to influence from wastewater effluent in the fall (Loomer 2008). Rainbow darter were collected using backpack electrofishing units (Smith Root LR 20/24) in riffle/run habitats with boulder and cobble substrate. All fish were collected and handled in accordance with the University of Waterloo's Animal Care Committee and Canadian Council on Animal Care Protocol (AUP # 40318). Approximately 40 male fish were collected from each site to meet sample size requirements for the various studied endpoints. Fish were transported in aerated buckets to a mobile laboratory, where they were processed (usually within one hour of capture to ensure consistent results) (Marjan et al. 2018). Fish were euthanized by a concussion and spinal severance. Length (± 0.1 cm) and body mass (± 0.001 g) were measured to calculate condition factor ($K = 100 \times [\text{body weight}/\text{length}^3]$). Liver and gonad were weighed (± 0.001 g) individually and used to calculate liver somatic index ($LSI = 100 \times [\text{liver weight}/\text{body weight}]$) and gonadosomatic index ($GSI = 100 \times [\text{gonad weight}/\text{body weight}]$), respectively. The condition and somatic index data are available but are not included in this thesis as they are being reported elsewhere. One lobe of the testes (or both lobes if the total weight of the tissue was less than 20 mg) was stored in an excess (~7 mL) of Medium 199 buffer (25 mM Hepes, 4.0 mM sodium bicarbonate, 0.01% streptomycin and 0.1% bovine serum albumin) on ice for *in vitro* steroid analysis. The other lobe was fixed in Davidson's solution (10% glycerol, 10% acetic acid, 20% formaldehyde, 30% ethanol, 30% water) for histology.

2.2.4 Effluent and river chemistry

Effluent discharge (effluent flow, volume) and chemistry (ammonia, nitrate) data were obtained from the Region of Waterloo (Pam Law and Dominika Celmer-Repin, Region of Waterloo, personal communication). Ammonia measurements were collected weekly from 2007 to 2013, three times per week from 2014 to 2018, and weekly in 2019. Nitrate measurements were collected approximately weekly from 2007 to 2019. Effluent volume and flow measurements were collected daily.

Effluent estrogenicity was measured in grab samples of the final effluent using the YES assay following methods outlined in Srikanthan (2019) and data were obtained from Srikanthan (2019).

Grand River flow and surface water temperature data for Bridgeport (station 68) and water temperature data for Below Shand Dam (station 88) were downloaded from the GRCA monitoring dataset made available under the Grand River Conservation Authority's Open Data Licence v2.0 on January 5, 2020 (Grand River Conservation Authority 2020). All GRCA data were measured hourly and are reported in this thesis as annual or monthly averages. Flow data for the Grand River at West Montrose (station 02GA034) were downloaded from the Environment and Climate Change Canada Historical Hydrometric dataset on March 19, 2020 (Water Survey of Canada 2020). Water Survey of Canada data were measured daily.

2.2.5 Stable isotope analysis

$\delta^{15}N$ and $\delta^{13}C$ in rainbow darter muscle tissue were analyzed at each site to compare isotopic signatures upstream and downstream of the Waterloo WWTP before and after its upgrades. Stable isotope data for 2007, 2008, and 2010–2014 were obtained from Hicks et al. (2017b). Methods for fish collected from 2016 to 2018 (n = 130) followed Loomer et al. (2015) and Hicks et al. (2017b). Briefly, 8–10 fish per site were selected within a range of 4.7–6.3 cm total length to limit variation between larger and smaller fish since the turnover rate of carbon is fastest in rapidly growing animals (Angradi 1994). These fish were not aged for this study but following growth rates defined in Crichton (2016), these fish would be designated in the 2+ age category. After field collection, fish bodies were stored at −20°C before a skinless piece of muscle was removed, dried at 60°C for 3–4 h or until completely dry, and ground into a powder. A sample of tissue (0.25–0.40 mg) was weighed into tin cups and analyzed using a 4010 Elemental Analyzer (Costech Instruments, Italy) coupled to a Delta Plus XL (Thermo Finnigan, Germany) continuous flow isotope ratio mass spectrometer (CFIRMS) to determine carbon and nitrogen composition. The $\delta^{13}C$ and $\delta^{15}N$ values are the corrected delta values, reported in per mil (‰) units, against the primary reference scale of Vienna Pee Dee Belemnite (VPDB) and atmospheric air, respectively. A subset (n = 24) of the samples analyzed for this study were run in duplicate and the mean ± SE differences between replicates were 0.05 ± 0.01‰ for $\delta^{13}C$ and 0.23 ± 0.04‰ for $\delta^{15}N$. C:N ratios were considered low (3.44 ± 0.01, n = 139) and no lipid normalization was necessary.

2.2.6 *In vitro* steroid production

Stimulated *in vitro* steroid analysis was used to analyze 11KT and testosterone steroid production, since it is difficult to obtain enough blood to study plasma steroids in small-bodied fish such as rainbow darter (McMaster et al. 1995). *In vitro* steroid levels have been found to correlate with circulating steroid levels

and are an accepted measure of the gonad's ability to produce steroids (McMaster et al. 1995). Since sampling time, capture stress, and confinement stress have been shown to alter sex steroid production in fish (Jardine et al. 1996), care was taken to sample fish at consistent times of day and the effect of holding times used (generally < 1 hr) was confirmed to not affect the results (Marjan et al. 2018). Therefore these factors were standardized among sites. In 2011, radioimmunoassays were used for steroid quantification (Fuzzen et al. 2016). From 2012 onwards, enzyme-linked immunosorbent assays (ELISAs) were conducted (Fuzzen et al. 2016; Marjan et al. 2018). Data from 2011 to 2016 were obtained from Fuzzen et al. (2016) and Marjan et al. (2018). Fish collected from 2017 to 2019 followed methods outlined in Marjan et al. (2018). Briefly, 10–20 mg of testes tissue from each fish (n = 18–20 per site) were placed in an excess of Medium 199. Samples were stored on ice during transport back to the laboratory at the University of Waterloo, where the gonad tissue weight was recorded, placed in 1 mL of fresh media, stimulated with 10 µL of 1 IU/mL human chorionic gonadotropin (hCG) dissolved in Medium 199, and incubated at 16°C for 24 h. After incubation, media was removed and stored at 80°C. Later, steroid analysis was conducted using an ELISA following manufacturer's (Cayman Chemical Company) instructions, with a range of 0.78–100 pg/mL for 11KT and 3.9–500 pg/mL for T. Samples were generally diluted ~40x–80x for 11KT and ~5x–10x for testosterone in order to fall within these ranges. Samples were run in triplicate and if the coefficient of variation (CV = (standard deviation/mean) × 100%) was greater than 20% between replicates then the sample was re-analyzed. If the sample concentration was outside the linear range of the standard curve (a standard curve was run on each plate), then the sample was re-analyzed at a different dilution. Inter-plate variation was controlled by using an internal standard: a pooled sample consisting of 3 samples per site that was analyzed on each plate. If the CV between internal standards was greater than 20% between plates then the plates were re-run. Blank Medium 199 spiked with hCG was also run in triplicate on each plate to confirm that Medium 199 was not interfering with the results of the immunoassay.

2.2.7 Intersex incidence and severity

Intersex incidence and severity data were obtained for 2007 and 2010–2015 from Hicks et al. (2017a). Histological methods for fish collected from 2016 to 2019 followed methods from Hicks et al. (2017a). Approximately 25 male fish were collected for intersex assessment from each site. After removal from the fish, one lobe of testis was placed in a histology cassette in Davidson's solution for 36–48 h before transferring to 70% ethanol until processing. Gonads were then dehydrated and embedded in paraffin wax using a tissue processor (Sakura Tissue-Tek VIP). Samples were sectioned at a thickness of 5 µm, placed

on slides, and stained with hematoxylin and eosin (Leica Autostainer XL, MRM Histology Laboratory, Canada Centre for Inland Waters, Burlington, ON). At least 40 sections per fish were analyzed for intersex at 100x magnification using a Leica DM100 light microscope. Intersex incidence for each site was determined based on the presence or absence of oocytes in the tissue and calculated by dividing the number of fish displaying intersex by the number of fish studied. Intersex severity was scored using the index developed by Bahamonde et al. (2015a) (Table 2.2).

Table 2.2 Intersex severity index (adapted from Bahamonde et al. 2015a).

Index	Criteria
0	100% male tissue.
1	Less than 1–3 primary oocytes per testis.
2	4–10 perinuclear oocytes per lobe.
3	>10 perinuclear oocytes.
4	Several sections with perinuclear and cortical alveolar oocytes. Clusters of eggs.
5	Less than 50% ovarian tissue, presence of vitellogenic eggs.
6	More than 50% ovarian tissue, presence of vitellogenic eggs.
7	100% female gonad.

2.2.8 Statistics

All analyses were conducted using R (version 4.0.0; R Core Team 2020) and graphs and data summaries were computed using the tidyverse packages (Wickham et al. 2019). Due to the unbalanced design, type III sum of squares (car package; Fox and Weisberg 2019) were used whenever an ANOVA was conducted. Sample size information for all biological endpoints are available in Appendix B1: Table S2.1.

Annual effluent nutrients (ammonia and nitrate) and river flow at West Montrose from 2007 to 2019 were analyzed following methods in Hicks et al. (2017a, 2017b), using a Kruskal–Wallis test with Dunn's post-hoc (using Holm–Bonferroni *p*-value correction; Holm 1979) and assessed at a significance level of alpha = 0.05.

Estrogenicity data were analyzed using the same methods as Srikanthan (2019). Data were assessed by comparing each sampling time point to the 95% confidence interval calculated from the pre-upgrade period which included samples collected in August, September and November 2015.

Stable isotope analysis followed methods published in Hicks et al. (2017b). To assess if there was a change in stable isotopes associated with the Waterloo WWTP upgrades, and to control for natural spatial variability moving downstream, a two-way ANOVA was conducted across years comparing DSW

and REF3 only. Tukey's post-hoc test was conducted on the interaction model when the interaction (year x site) was significant and on the additive model when the interaction was not significant. To assess spatial changes across the sites, one-way ANOVAs were conducted for each year comparing all sites. By testing each year individually, this method controlled for natural variability among years. Assumptions for ANOVA (normality of residuals and equal variance) were often not met even with transformation, and therefore significance was assessed at a more conservative alpha threshold of 0.01 (Pitt et al. 2009; Hicks et al. 2017b). Pairwise comparisons were conducted using Tukey's post-hoc test and significance was also assessed at alpha = 0.01.

To test if there was a relationship between effluent ammonia concentrations and $\delta^{15}N$ signatures in rainbow darters living downstream of the WWTP, a Pearson correlation was conducted on the mean effluent ammonia over the two-month period prior to sampling (September–October) and the mean $\delta^{15}N$ value for fish caught at DSW. A Pearson correlation was also conducted between the mean two-month river flow (September–October) and mean $\delta^{13}C$ at one of the reference sites, REF2. REF2 was chosen since it does not have influence from the Shand Dam (Loomer 2008) and because there is a flow gauge (Water Survey of Canada 2020) at that location. A two-month time period was chosen because the half-life of $\delta^{13}C$ and $\delta^{15}N$ turnover in rainbow darter muscle tissue was determined to be 33 and 29 days, respectively (Hicks et al. 2017b), and the two-month period best described the variation in isotopic ratios.

In vitro 11KT and T production were analyzed using a combination of methods previously published (Fuzzen et al. 2016; Marjan et al. 2018). All steroid data were log10 transformed to meet assumptions of normality and homogeneity of variance. Since not all sites were sampled in every year, running a two-way ANOVA was not possible, so each year was analyzed individually using a one-way ANOVA (Marjan et al. 2018) with Tukey's post-hoc test (alpha = 0.05). To put the statistics into context and help visualize site differences, the 95% confidence interval and the 25% critical effect size (Munkittrick et al. 2009) calculated from the pooled data from REF2 was added to the graphs (Fuzzen et al. 2016). REF2 was chosen because it was hypothesized to be least affected by confounding environmental factors and visually had the most consistent steroid production across years. To investigate if there were differences in steroid production at DSW among years and to account for yearly variability and differences in methodology between 2011(radioimmunoassay) and 2012–2019 (ELISA), the logged data were normalized to the steroid production at REF2 within each year, and then compared using a one-way ANOVA with Tukey's post-hoc test (alpha = 0.05).

Intersex incidence and severity were analyzed using methods from Hicks et al. (2017a). Because there were not enough post-upgrade years sampled to conduct a before-after-control-impact (BACI)

analysis, each site was compared across years and each year was compared across sites in separate analyses. Incidence data were analyzed using Fisher's exact test and with Fisher's exact pairwise comparisons post-hoc test (rcompanion package; Mangiafico 2020) with a Holm–Bonferroni *p*-value correction (Holm 1979). Severity score data were analyzed using a Kruskal–Wallis test and Dunn's post-hoc with a Holm–Bonferroni *p*-value correction (Holm 1979). Significance was assessed at alpha = 0.05. A Pearson correlation was conducted to investigate the relationship between intersex incidence and severity and included all sites and all years.

2.3 Results

2.3.1 Water and effluent characteristics

2.3.1.1 Effluent nutrients

Effluent chemistry data collected from the Waterloo WWTP indicated that effluent quality changed over time during the upgrades. In 2007 and 2008, ammonia concentrations in the effluent were relatively low (median annual value of 6.7 mg/L for 2007 and 2008 combined; Figure 2.2A). From 2009 to 2014 there were construction delays at the plant which resulted in partial nitrification. Ammonia concentrations began to increase in 2009 and were at their highest median concentrations (25–26 mg/L) in 2012 and 2013. Between the years 2009 to 2015, ammonia concentrations remained significantly higher than they were in 2007 and 2008. In 2016, effluent quality began to improve, and the median ammonia concentration dropped to 12.7 mg/L. In the springs of 2017 and 2018 the two aeration tanks were upgraded, and ammonia continued to decrease (2017 = 4.8 mg/L; 2018 = 2.5 mg/L; 2019 = 0.4 mg/L). Nitrate concentrations in the effluent were consistently low from 2007 to 2011 (median concentration 6.5 mg/L for the five years combined; Figure 2.2B). In 2012 and 2013, nitrate concentrations in the effluent were even lower (median 0.8 mg/L for the two years combined). In 2014, nitrate concentrations began to increase and reached their highest median value of 33.3 mg/L in 2018 and 2019.

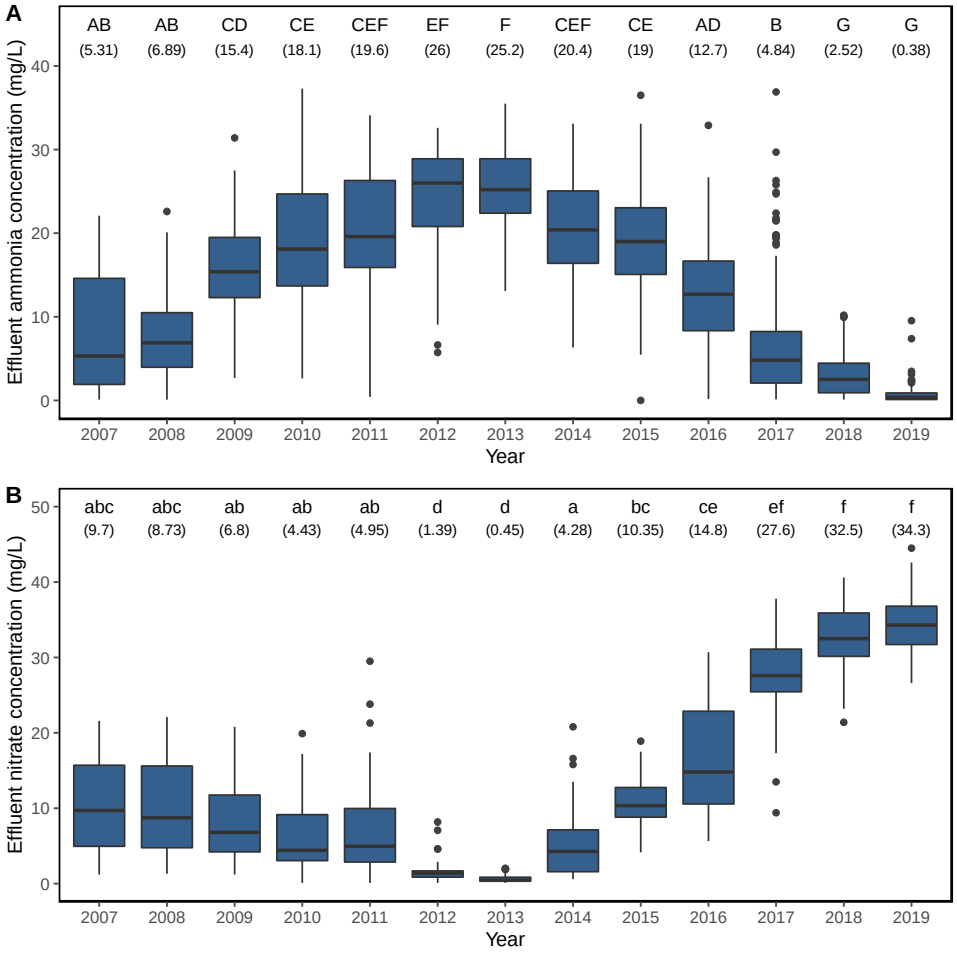

Figure 2.2 Boxplots of (A) ammonia and (B) nitrate concentrations in the final effluent of the Waterloo WWTP. The bars indicate the median value, the bottom and top of the box are the first and third quartiles, and the whiskers extend to the largest value not farther than 1.5 × the interquartile range. Outliers are plotted individually. Numbers in brackets indicate the median nutrient concentration for that year. Years not sharing a letter are significantly different.

2.3.1.2 Effluent estrogenicity

Estrogenicity in the Waterloo WWTP effluent was highly variable. Before the upgrades were completed, estrogenicity was generally 5–15 ng/L E2eq, but there were spikes up to 32 ng/L (Figure 2.3). After the upgrades were completed, estrogenicity fell below the 95% confidence interval (range of 4.1–12.8 ng/L E2eq) for the pre-upgrade mean. Concentrations were < 1 ng/L E2eq by November 2018.

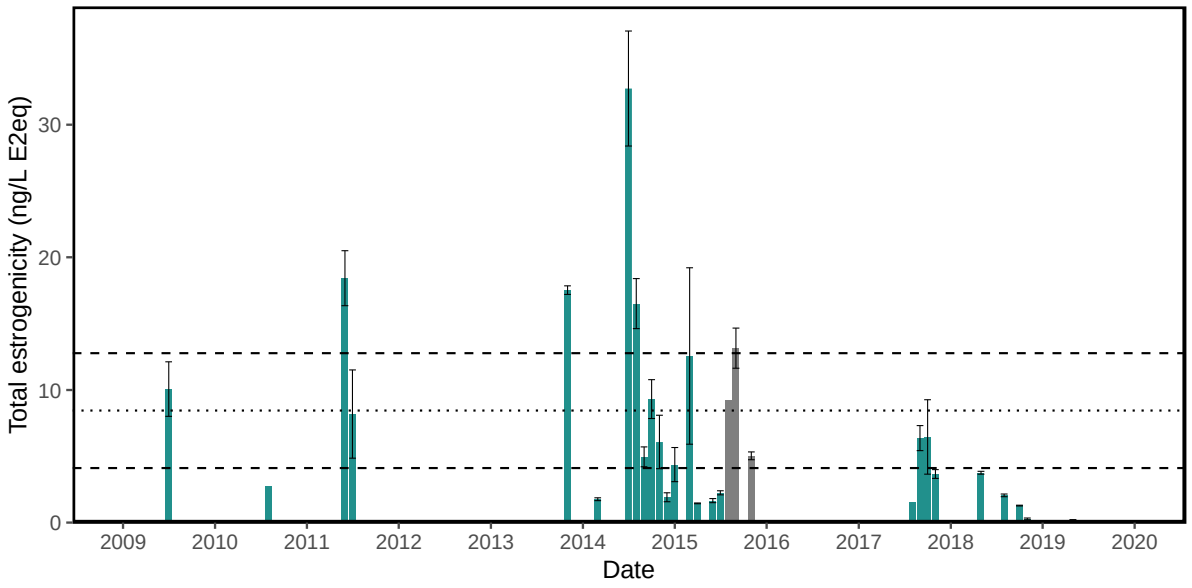

Figure 2.3 Mean effluent estrogenicity (± SE) collected over the course of the Waterloo WWTP upgrades. The dotted line is the pre-upgrade mean calculated using the data from August, September, and November 2015 (grey bars), and the dashed lines indicate the 95% confidence interval of this mean. Data were from Srikanthan (2019) and graph is modified from Srikanthan (2019) to include additional data collected after May 2018.

2.3.1.3 River flows

River flows were variable and low flow occurred in mid-summer each year. Annual flows were lowest in 2007 and 2016. The highest flows were in 2008, 2013, and 2017 (Kruskal–Wallis test with Dunn's post-hoc, $p < 0.05$; Figure 2.4). There was a drought in the late summer of 2016. 2017 was a wet year in comparison.

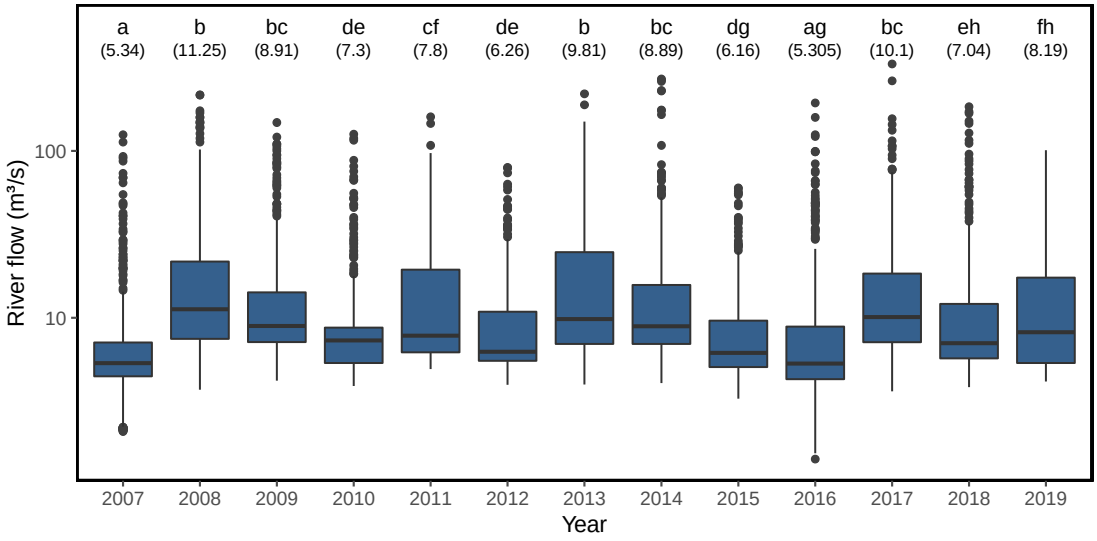

Figure 2.4 Boxplots of annual river flow (m^3/s) at REF2. The bars indicate the median value, the bottom and top of the box are the first and third quartiles, and the whiskers extend to the largest value not farther than 1.5 × the interquartile range. Outliers are plotted individually. Note the y-axis is plotted on a log10 scale for better visualization. Numbers in brackets indicate the median river flow for each year. Years not sharing a letter are significantly different.

2.3.2 Stable isotopes

2.3.2.1 Stable isotope ratio of $\delta^{13}C$

There were differences in $\delta^{13}C$ among sites for all 10 years studied (one-way ANOVAs, $p < 0.01$). In general, $\delta^{13}C$ signatures became more enriched moving downstream (Figure 2.5A). The farthest upstream reference site, REF1, was generally not statistically different from the next site downstream (except in 2012 when it was more enriched). REF2 always had the lowest $\delta^{13}C$ signature. REF3 was always an intermediate between REF2 and DSW; it was significantly lower than DSW in some years (2008, 2011, 2014, 2016, 2017, 2018), significantly higher than REF2 in one year (2012) and not statistically different from either site in two years (2007, 2013). $\delta^{13}C$ at DSW was always significantly more enriched than at REF2. $\delta^{13}C$ at INT1 had a more variable response and did not follow the pattern of downstream enrichment. It was either not statistically different from DSW (2014, 2016, 2017, 2018) or significantly lower (2013). Full results from one-way ANOVAs are available in Appendix B2: Table S2.2.

The stable isotope ratio of $\delta^{13}C$ was significantly different between DSW and REF3 (two-way ANOVA; $F_{1,146} = 119.2$, $p < 0.001$) and among years ($F_{8,146} = 16.38$, $p < 0.001$) but there was no significant interaction between site and year. On average, the stable isotope ratio of $\delta^{13}C$ at DSW was 1.12‰ higher than at REF3 (Tukey's post-hoc test, $p < 0.001$). Average $\delta^{13}C$ varied among years and was lowest in 2014 and highest in 2016. Full results from two-way ANOVAs are available in Appendix B2: Table S2.3.

There was a negative correlation between mean river flow calculated during the two months prior to sampling and the mean $\delta^{13}C$ ratio in rainbow darter caught at the reference site REF2 (Figure 2.5B; Pearson's $r = -0.81$, $p = 0.014$).

2.3.2.2 Stable isotope ratio of $\delta^{15}N$

$\delta^{15}N$ tended to become more enriched moving downstream, except for REF1 which was not significantly different from REF2 (Figure 2.5C). There was less annual variability in $\delta^{15}N$, but $\delta^{15}N$ signatures at DSW were low during the years when the Waterloo WWTP was releasing poor quality effluent (2011–2014). $\delta^{15}N$ exhibited significant differences among sites for each year tested except 2007 and 2008 (one-way ANOVAs, $p < 0.001$). The three reference sites generally had similar $\delta^{15}N$ values. They were not significantly different from each other in 2012, 2013, or 2014. REF1 was never significantly different from REF2, though it was sometimes an intermediary between REF2 and REF3, with REF3 being higher than REF2 in 2011, 2016, 2017, and 2018. $\delta^{15}N$ values at DSW varied greatly. During the years of poor effluent quality, DSW had the lowest $\delta^{15}N$ (2011–2014). As the effluent quality improved, $\delta^{15}N$ at DSW became similar to (2016) or higher than (2017, 2018) REF3. $\delta^{15}N$ at INT1 was higher than DSW in most years studied (2013, 2014, 2016, 2017) except in 2018 when the two sites were not significantly different. Full results from one-way ANOVAs are available in Appendix B2: Table S2.2.

There was a significant interaction for $\delta^{15}N$ between DSW and REF3 among years (two-way ANOVA, $F_{8,140} = 21.0$, $p < 0.001$) but no main effects of site ($F_{1,140} = 0.62$, $p = 0.43$) or year ($F_{8,140} = 1.62$, $p = 0.12$). In 2007 and 2008 before the upgrades, $\delta^{15}N$ was not different between REF3 and DSW. During the upgrades from 2011 to 2014, $\delta^{15}N$ was significantly lower at DSW by 2.59–3.49‰ ($p < 0.001$). In 2016 and 2017 as effluent quality was improving, there were again no differences between REF3 and DSW. In 2018, $\delta^{15}N$ at DSW was higher than REF3 by 3.98‰ ($p < 0.001$). There were no significant differences in $\delta^{15}N$ at REF3 among years. Full results from two-way ANOVAs are available in Appendix B2: Table S2.4.

$\delta^{15}N$ stable isotope signatures had a negative relationship with mean effluent ammonia concentrations calculated based on the two months prior to fish sampling date (Figure 2.5D; Pearson's $r = -0.94$, $p < 0.001$).

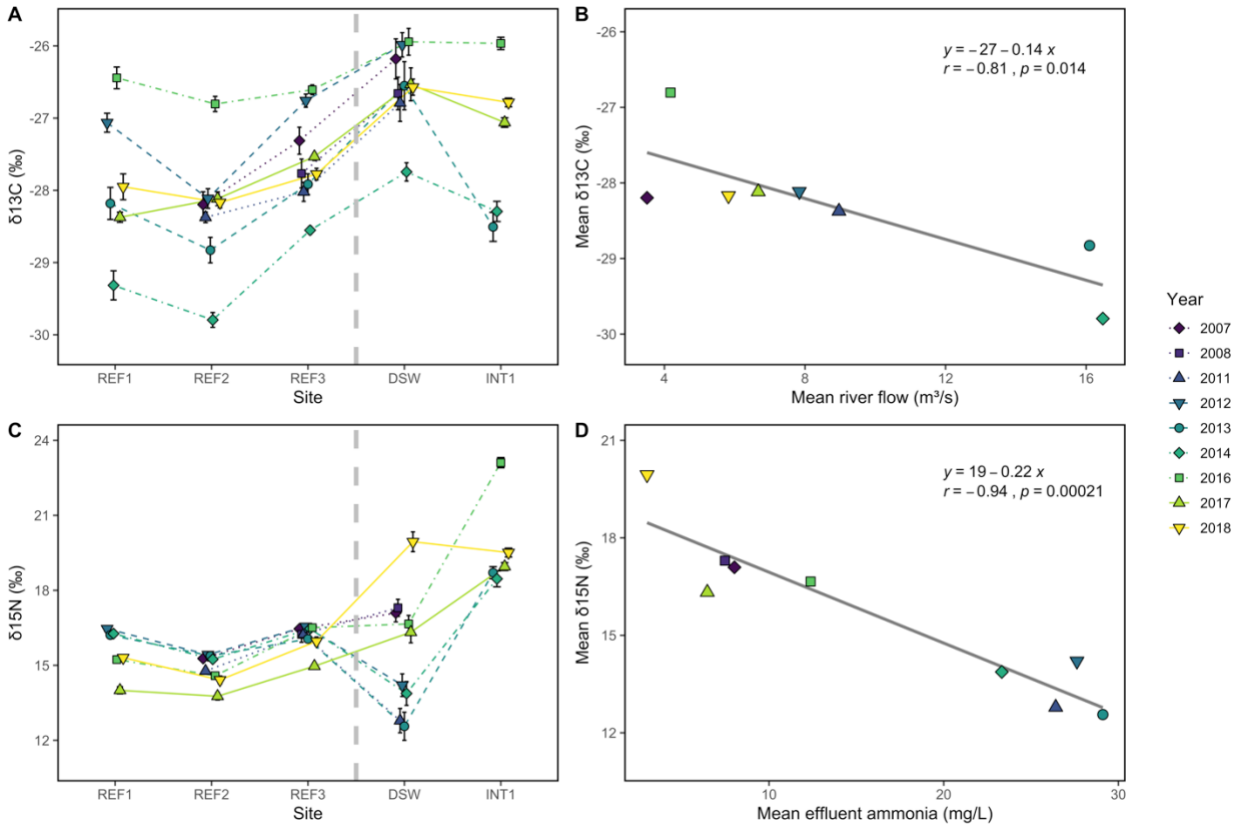

Figure 2.5 Stable isotope ratios (mean ± SE) in rainbow darter in the Grand River. (A) $\delta^{13}C$ isotopic ratios in rainbow darter in the Grand River at sites upstream and downstream of the Waterloo WWTP (location indicated by the dashed line). (B) Average $\delta^{13}C$ was strongly correlated with average flow in the two months prior to sampling at reference site REF2 (note $\delta^{13}C$ is NA for REF2 in 2008). (C) $\delta^{15}N$ isotope ratios in rainbow darter caught at sites upstream and downstream of the Waterloo WWTP (location indicated by dashed line) in the Grand River. (D) Average $\delta^{15}N$ in fish tissue was correlated with the average ammonia concentrations in the effluent downstream of the Waterloo WWTP.

2.3.3 *In vitro* steroid production

2.3.3.1 Spatial patterns in 11KT and T production

In vitro production of 11-ketotestosterone (11KT) and testosterone (T) varied greatly among sites and years, with few consistent patterns. If there were differences in steroid production among sites, DSW tended to have lower production compared to at least one of the three reference sites, but this pattern was slightly different each year and for each steroid tested (Figure 2.6).

2.3.3.2 11KT before and after upgrades

Different patterns were observed in 11KT production when comparing DSW with each of the three reference sites. In three of the pre-upgrade years, DSW 11KT production was not statistically different from any of the reference sites (2011, 2014, 2016). In one year (2013), DSW 11KT production was significantly lower than REF1 and REF2 but not REF3. DSW 11KT production was significantly lower than all three reference sites in 2012 and 2015. 11KT production at DSW was lower than the 25% critical effect size (CES) for REF2 in 2012, 2013, and 2015. 11KT production in rainbow darter caught farther downstream at INT1 also varied among years but was never lower than DSW. In 2017, the year of the first aeration tank upgrades, DSW 11KT production was significantly lower than REF1 and REF2 but not REF3. In 2018, the first full season after both aeration tank upgrades were completed, 11KT at DSW was lower than REF3, but there were no other differences among sites. In 2019 there were no significant differences among sites (one-way ANOVAs with Tukey's post-hoc test). Full results are available in Appendix B3: Table S2.5.

2.3.3.3 T before and after upgrades

Testosterone production was also variable among sites. In two pre-upgrade years (2011, 2014), DSW T production was not different from any of the three reference sites. In one year (2012), DSW T production was significantly lower than REF2 and REF3 but not REF1. In one year (2013), DSW was lower than only REF2. T production was lower than all three reference sites in 2015 and 2016. Despite these statistical differences, T production at DSW was only lower than the 25% CES for REF2 in 2012 and 2014. In fall 2017 after the first set of aeration tank upgrades, T production at DSW was lower than one reference site: REF1. In 2018, the first full year after the upgrades, there was variable T production: DSW was lower than REF1, REF3, and INT1 but there were no other differences among sites and it was not lower than the 25% CES threshold. In 2019, there were no differences among sites (one-way ANOVAs with Tukey's post-hoc test). Full results are available in Appendix B3: Table S2.5.

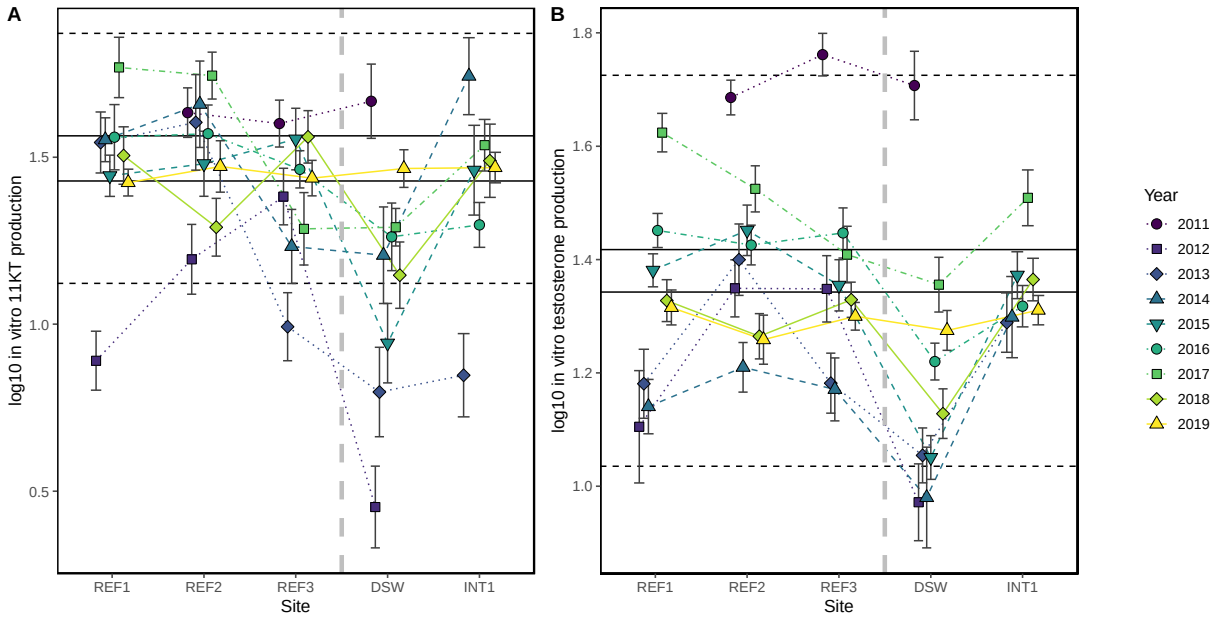

Figure 2.6 Log10 transformed *in vitro* steroid production (mean ± SE) of (A) 11KT and (B) testosterone at each site. For both A and B, the horizontal dashed lines indicate the 25% critical effect size (CES) calculated from the mean of the data from REF2, and the horizontal solid lines indicate the 95% confidence interval calculated from the data from REF2. The vertical gray dashed line indicates the location of the Waterloo WWTP.

2.3.3.4 Yearly patterns in steroid production at DSW

DSW steroid production varied greatly among years. To analyze these responses and control for yearly variability, data were logged to meet assumptions of ANOVA and normalized to the average steroid production at REF2 for each year (Figure 2.7). Both 11KT and T production at DSW were highest in 2011 (11KT production was 102% of the production at REF2 and T production was 101% of the production at REF2), dropped to their lowest values in 2012 (11KT production was 38% of the production at REF2 and T production was 72% of the production at REF2), and in general increased from 2013 to 2019 back to the same level as 2011 (one-way ANOVAs with Tukey's post-hoc test). Full results are available in Appendix B3: Table S2.6. 11KT at DSW was lower than the 25% CES for REF2 in 2012, 2013, and 2015. T was lower than the 25% CES only in 2012 and 2015. However, steroid production at

REF3 (grey diamonds in Figure 2.7) was often similar to DSW (i.e., lower than REF2), which confounds the interpretation of the relationship between DSW and REF2.

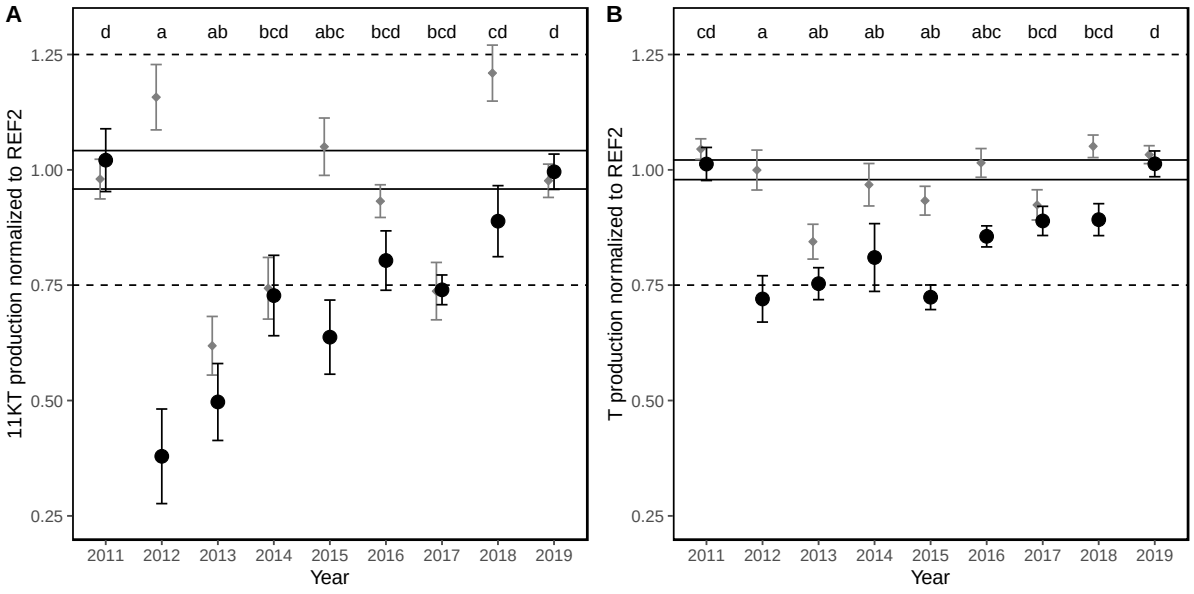

Figure 2.7 Mean (± SE) (A) 11KT and (B) T production at DSW (black circles), logged and normalized to the steroid production at REF2 for each year. Dissimilar letters indicate significant differences between years. The grey diamonds indicate steroid production at REF3. These points were added to the plot but not analyzed statistically. The horizontal dashed lines indicate the 25% critical effect size (CES) calculated from the mean of the data from REF2, and the horizontal solid lines indicate the 95% confidence interval calculated from the data from REF2.

2.3.4 Intersex

The highest level of intersex observed was a score of 4 (severe intersex). This score was observed in nine out of 878 fish studied from 2007 to 2019. Of those nine fish, seven were caught at DSW, one was caught at INT1, and one was from REF3. An example of a normal male rainbow darter testis and a male testis displaying intersex is shown in Figure 2.8.

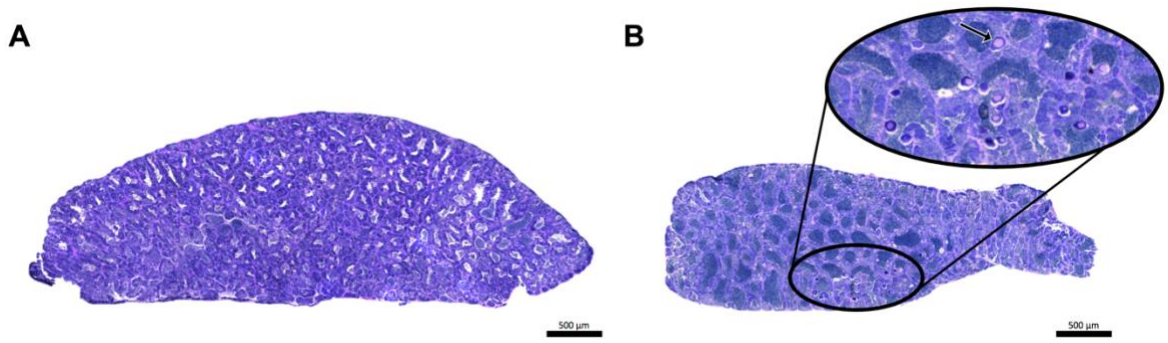

Figure 2.8 Micrographs of male rainbow darter testes. (A) Normal male rainbow darter testis (intersex score = 0). (B) Male rainbow darter testis with intersex (intersex score = 4). Some perinuclear oocytes are visible within the inset (one is indicated by the arrow). Scale bars are 500 µm.

2.3.4.1 Intersex incidence and severity before and after the upgrades

There were no differences in intersex incidence among years at any of the three reference sites or at the second downstream site, INT1 (Fisher's exact test, $p > 0.05$; full results available in Appendix B4: Table S2.7). Across all years, the portion of fish displaying intersex at REF1 was 8.5% (range 0% to 17%), at REF2 was 7.7% (range 0% to 20%), and at REF3 was 7.4% (range 0% to 17%). There were differences in intersex incidence at DSW among years (Fisher's exact test, $p = 0.013$), but no pairwise comparisons were significant after adjusting for multiple comparisons. The overall incidence of intersex at DSW was 31% (range 10% to 55%). Incidence at INT1 was 20.6% (range 12% to 33%). Intersex incidence across all sites over time is shown in Figure 2.9A.

There were also no differences in intersex severity among years at any of the three reference sites or the second downstream site (INT1), over the course of the study (Kruskal–Wallis test, $p > 0.05$; full results available in Appendix B4: Table S2.7). There were differences in intersex severity at DSW (Kruskal–Wallis test, H = 23.23, df = 9, $p < 0.01$), but only for the pairs 2015–2016 (Dunn's post-hoc, $p = 0.04$), 2016–2018 ($p = 0.047$) and 2016–2019 ($p = 0.006$). Intersex severity scores across all sites over time are shown in Figure 2.9B.

There were significant differences in intersex incidence among sites for some of the pre-upgrade years: 2011 (Fisher's exact test, $p = 0.02$), 2015 ($p = 0.03$), and 2016 ($p < 0.001$), as well as the transitional year 2017 ($p = 0.005$) and the post-upgrade year 2018 ($p = 0.03$). However, after adjusting for multiple comparisons, there were no significant pairwise differences among sites in 2011, 2015, and

2018. In 2016, incidence at DSW was significantly higher than REF1 (p = 0.01) and REF2 (p = 0001), and in 2017, incidence at DSW was significantly higher than REF3 (p = 0.01). There were no differences in intersex incidence among sites in 2007, 2012, 2013, 2014, and 2019. Full results are available in Appendix B4: Table S2.8.

Before the WWTP upgrades, there were significant differences in intersex severity among sites in 2011 (Kruskal–Wallis test, H = 7.87, df = 2, p = 0.02), 2014 (H = 9.75, df = 4, p = 0.04), and 2016 (H = 25.17, df = 4, p < 0.001). There were also differences among sites in fall 2017 after the first aeration tank was upgraded (H = 14.85, df = 4, p = 0.005). These differences were always between DSW and one or more of the reference sites: in 2011, DSW had a higher severity than REF2, in 2014 it was higher than REF1, in 2016 it was higher than all three reference sites, and in 2017 it was higher than REF3. There were no differences in intersex severity among sites in 2007 when the WWTP was operating well, or for some of the years during the upgrades (2012, 2013, and 2015) as well as the post-upgrade years 2018 and 2019. Full results are available in Appendix B4: Table S2.8.

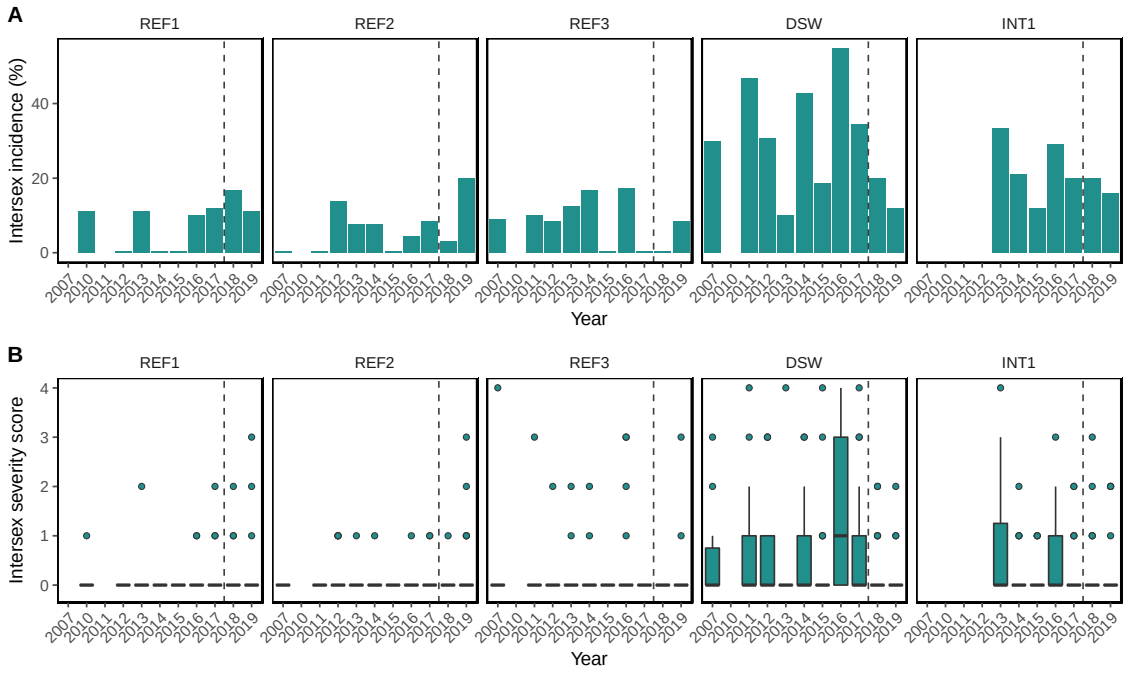

Figure 2.9 (A) Intersex incidence at each site over time. (B) Boxplot of intersex severity scores at each site over time. The bars indicate the median value, the bottom and top of the box are the first and third quartiles, and the whiskers extend to the largest value not farther than 1.5 × the interquartile range. Dashed vertical lines indicate the time of the WWTP upgrades.

Intersex severity was positively correlated with intersex incidence across all sites and years (Figure 2.10; Pearson's $r = 0.94$, $p < 0.001$).

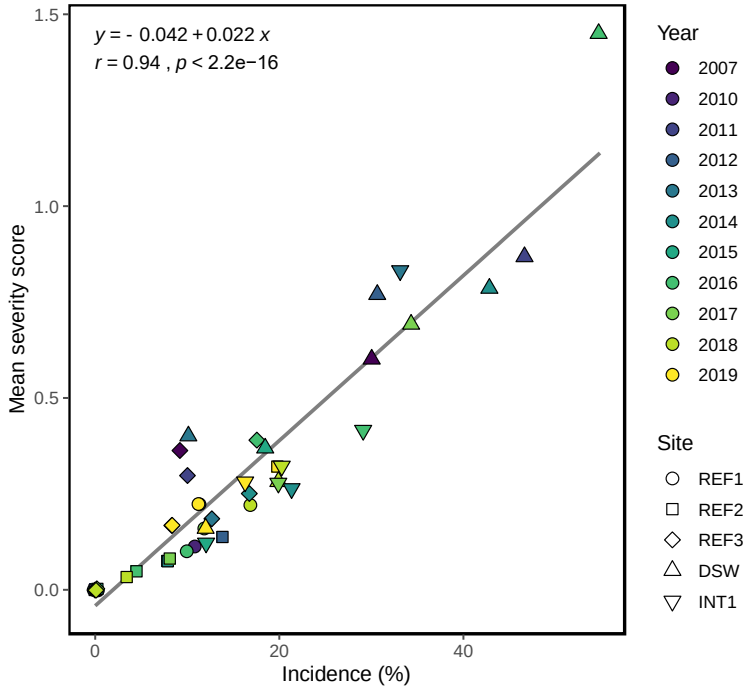

Figure 2.10 Correlation between mean severity score and intersex incidence (%) for each site and year (Pearson's $r = 0.94$, $p < 0.001$).

2.4 Discussion

Wastewater treatment plant upgrades are predicted to reduce the exposure of downstream fish to a variety of nutrients and contaminants, including EDCs. The Waterloo WWTP underwent major upgrades from 2009 to 2018. During this time effluent quality varied considerably as construction progressed. Effluent quality initially worsened for several years before improving once the upgrades were completed with the commission of one aeration tank in 2017 and a second aeration tank in 2018. The upgrades increased aeration and nitrification which resulted in lower ammonia and higher nitrate concentrations in the final effluent. There was also a reduction in selected pharmaceuticals and personal care products measured and a reduction in effluent estrogenicity after the upgrades were completed (Srikanthan 2019). Environmental exposure to wastewater in river systems can be highly modulated by flow and depending on the nature of the endpoint, this can influence how effects are manifested and interpreted in exposed fish populations.

Although the volume of wastewater released by the Waterloo WWTP was relatively constant, fish exposure to wastewater effluent below the Waterloo outfall varied over time because changes in river flows altered the dilution of the effluent in river water (Figure S1.2 and Figure S1.3). At its outfall, the Waterloo WWTP was estimated to contribute 2–5% of the volume of water in the Grand River assuming instantaneous mixing of the effluent and river water. However, since the effluent and river water are not fully mixed by the time it flows past DSW (1 km downstream; personal observations), fish at that site are likely exposed to much higher concentrations of effluent. Wastewater upgrades led to improvements in several biological endpoints (e.g., stable isotopes, intersex) in fish but it can be difficult to separate the impacts of wastewater from the upstream stressors and annual variability (e.g., *in vitro* sex steroid production).

2.4.1 Stable isotopes

Stable nitrogen isotope signatures in rainbow darter effectively tracked the changes in effluent quality (e.g., ammonia concentrations) over the course of the WWP upgrades. In contrast, stable carbon isotope signatures were found to correlate to changes in river flow but were not influenced by the WWTP upgrades. These results suggest that changes in stable isotope signatures in rainbow darter can give insight into effluent exposure and quality ($\delta^{15}N$) as well as responses to annual variation in flow ($\delta^{13}C$).

Stable carbon isotope signatures generally became more enriched moving downstream and this was consistent among years independent of the state of the WWTP upgrades. These results follow trends reported by Hicks et al. (2017b) after the Kitchener WWTP upgraded. Fish caught at REF1 were an exception to this pattern as they had slightly higher $\delta^{13}C$ values than expected, likely because of the site's proximity to the Shand Dam located approximately 18 km upstream. The Shand Dam is a bottom-draw dam which releases organic matter from its hypolimnion that is more enriched in $\delta^{13}C$ (Angradi 1994; Loomer 2008; Hicks et al. 2017b). There was always isotopic enrichment between the sites REF2 and DSW and usually enrichment between REF3 and DSW, which may be partially associated with the assimilation of sewage-derived nutrients but may also be attributed to the downstream trend of carbon enrichment (France 1995a; Finlay 2004; Loomer et al. 2015). However, farther downstream at INT1, $\delta^{13}C$ isotopic ratios were slightly lower, suggesting that there may have been an influence of the treatment plant or other inputs in the 11 km between the two sites. Downstream enrichment of $\delta^{13}C$ has been reported in other river systems (Rasmussen and Trudeau 2007). Enrichment of $\delta^{13}C$ is often found in relation to the assimilation of wastewater-derived nutrients (Brown et al. 2011; Loomer et al. 2015; Robinson et al. 2016; Hicks et al. 2017b), but some studies report depletion (deBruyn and Rasmussen

2002) and other studies did not see any changes in $\delta^{13}C$ (Morrissey et al. 2013). These differences may be specific to the study system, receiving water body, type of wastewater treatment, and the dynamics and flow of the effluent plume.

$\delta^{13}C$ signatures had a negative relationship with average river flows which explained the variability in $\delta^{13}C$ among years. This relationship was also demonstrated previously in the Grand River (Hicks et al. 2017b). In the current study, the two-month period prior to sampling had the best relationship with $\delta^{13}C$, which lends support to the reported 33 day half-life of $\delta^{13}C$ turnover in rainbow darter muscle tissue (Hicks et al. 2017b), and also supports the general idea that $\delta^{13}C$ reflects an organism's diet assimilated over weeks or months (Angradi 1994). Many studies have found a similar negative relationship between river flow and $\delta^{13}C$ signatures in primary producers (Finlay et al. 1999; Rasmussen and Trudeau 2007). Less fractionation of carbon occurs at low water velocities because of the stagnant boundary layer which increases resistance to diffusion, and this results in signatures that are more enriched in ^{13}C (France 1995b; Rasmussen and Trudeau 2007, 2010). These carbon isotopes then are transferred up the food web to fish consumers (Rasmussen and Trudeau 2010).

Stable nitrogen isotope signatures in rainbow darter closely tracked the upgrade status and effluent quality of the Waterloo WWTP. There was a clear difference between $\delta^{15}N$ in years with poor quality effluent (i.e., high ammonia) and the years with higher quality effluent. During the WWTP construction when the Waterloo WWTP was releasing poor-quality effluent, $\delta^{15}N$ signatures at DSW were much lower than they were at the reference sites. Then, once nitrification of the effluent improved, $\delta^{15}N$ signatures became more enriched and resembled upstream values. This response followed the same pattern that was observed after the Kitchener WWTP upgraded in 2012, with $\delta^{15}N$ enrichment after the upgrades (Hicks et al. 2017b). In another study, $\delta^{15}N$ signatures in algae and crabs became more depleted following WWTP upgrades, but this was likely driven by the biological nutrient removal process employed which dramatically reduced N loadings into a marine system (Pitt et al. 2009). Regardless, each study showed $\delta^{15}N$ values resembling reference site levels after their respective WWTP upgrades.

There was a strong negative relationship between $\delta^{15}N$ in rainbow darter and average effluent ammonia over the two-month period prior to sampling. This finding supports the similar relationship originally reported with data from both the Waterloo and Kitchener WWTPs (Hicks et al. 2017b). During nitrification, nitrogen in the form of urea is converted to ammonium and then nitrate (Heaton 1986). The light isotope of nitrogen is preferentially lost during this process, leaving behind ammonium and nitrate that is enriched in ^{15}N (Heaton 1986). $\delta^{15}N$ has commonly been used to assess anthropogenic inputs to aquatic ecosystems (Cabana and Rasmussen 1996; Anderson and Cabana 2006). Enriched $\delta^{15}N$ signatures

in aquatic food webs have been used as a tool to assess the location of sewage and wastewater inputs (Steffy and Kilham 2004; Morrissey et al. 2013). Anderson and Cabana (2006) found strong positive correlations between anthropogenic sources of N and $\delta^{15}N$ in primary consumers, with the strongest relationship driven by livestock, so it was surprising that the $\delta^{15}N$ signatures at our sites downstream of the highly agricultural Conestogo River did not seem to be greatly affected (Appendix A2: Figure S1.5).

Overall, stable isotope signatures seem to be a robust tool for assessing recent exposure to wastewater effluent. In the Grand River, rainbow darter have small home ranges (Hicks and Servos 2017) and their stable isotopes signatures appear to represent site-specific changes. Additionally, the half-lives for ^{13}C and ^{15}N isotope turnover in rainbow darter muscle tissue are 33 and 29 days respectively (Hicks et al. 2017b) and therefore their isotopic signatures reflect their somewhat recent (weeks to months) diet. Loomer et al. (2015) showed that $\delta^{15}N$ signatures change seasonally but that their relationship to wastewater outfalls remained consistent. Our results showed that $\delta^{15}N$ responded to changes in effluent ammonia, and therefore supports the use of $\delta^{15}N$ as an indicator of effluent exposure in fish during WWTP upgrades.

2.4.2 *In vitro* steroid production

In vitro steroid production of 11KT and T was highly variable among sites and years and therefore it was difficult to link to changes in effluent quality. Before the upgrades were finished, 11KT production downstream of the Waterloo WWTP was lower than one or more reference sites in only three of the six years studied. T was lower than one or more reference sites in four of the six years. 11KT production at DSW fell below the 25% critical effect size (CES) threshold for all of those three years, while T production at DSW was only lower than the 25% CES for two of the years. In the fall after the first aeration tank upgrades were completed (2017), there was a depression of 11KT production at DSW compared to REF2 and a depression of T production at DSW compared to REF1, but neither 11KT or T production was lower than the 25% CES threshold. Steroid production was variable but not lower than the 25% CES in 2018, and in 2019 there were no differences in steroid production among sites. Recovery of *in vitro* steroid production was gradual after the Kitchener WWTP was upgraded (Marjan et al. 2018), but it was difficult to make firm conclusions given the variability in the Waterloo data, even though steroid levels downstream of the Waterloo WWTP now resemble the reference sites (in 2019).

Although there are no longer reductions in steroid production downstream of the WWTP, there are still unknown factors that contributed to variable steroid production throughout the study period and may still have the potential to affect steroid production in the future. For example, in many years there

was a depression in steroid production at REF3 which masked the effect at DSW. REF3 is located downstream of the confluence of the Conestogo River, a highly agricultural tributary of the Grand River (Loomer and Cooke 2011). To investigate if this had an effect on steroid production at REF3, a site was studied in the Conestogo River and upstream and downstream of its confluence with the Grand River in 2015 (Marjan et al. 2018) and 2018 (Appendix A2: Figure S1.6). However, no reduction of steroid production was found at REF3 in either 2015 or 2018, and therefore there was no evidence that the Conestogo River was causing the depression in steroid production at REF3 that had been found previously. The effect seems to be transient and only occurs in some years. REF3 is also located within the City of Waterloo, so there may also be influence associated with urbanization or other anthropogenic stressors. For example, metal concentrations in mussel gills have been shown to be increased at urban sites (Gillis 2012). Runoff from impervious surfaces can introduce various contaminants and increase river temperatures (Nelson and Palmer 2007; Kaushal et al. 2010). The interaction of temperature with photoperiod is known to affect steroid production in teleosts (de Vlaming 1972). Water temperatures have been shown to increase moving downstream in the Grand River due to the influence of urban runoff and WWTP effluent (Hicks 2017), and changes to flow and water levels affect water temperature as well (Kaushal et al. 2010). Changes in flow can also affect the dilution and movement of the effluent plume or other runoff within the river, altering fish exposure to contaminants (Marjan et al. 2017). However, there was no correlation between 11KT or T and river flow (Appendix A4: Figure S1.12). These interacting environmental conditions may have affected the timing of the recrudescence period at some sites differently than others or contributed to the apparent gradient of steroid depression moving downstream.

Wild fish caught downstream of WWTP effluents have been shown to have lower production of sex steroids and these effects have been attributed to estrogenic compounds in the effluent (Folmar et al. 1996; Weber et al. 2019). Fish living downstream of combination of inputs from WWTPs, agriculture, and chemical plants have been shown to exhibit lower production of 11KT and variable production of T (Hecker et al. 2002). Laboratory exposures have confirmed that exposure to environmentally relevant concentrations of EE2 (10 ng/L; Salierno and Kane 2009) and E2 (50 ng/L; Martinović et al. 2007) can suppress sex steroid production in fathead minnow. Within the Grand River, effluent from the Kitchener WWTP was associated with reductions in sex steroid production in rainbow darter which then improved after the WWTP was upgraded (Fuzzen et al. 2016; Marjan et al. 2018). The upgrades to the Kitchener WWTP lowered the estrogenicity in the effluent (Srikanthan 2019), and therefore it is certainly plausible that exogenous EE2 plays a large role in suppressing androgen production in male rainbow darter. The Waterloo WWTP effluent had variable estrogenicity during its upgrades and estrogenicity decreased after

the upgrades were completed (Srikanthan 2019), but there was no correlation between effluent estrogenicity and steroid response (Appendix A4: Figure S1.13). However, estrogen sampling did not occur in every year and not always at times corresponding to fish sampling (or their sensitive windows that may influence steroid production), and therefore it remains possible that variable estrogenicity contributed to the variable steroid response depending on the time of year as well as how long the fish were exposed.

It is also possible that fish living downstream of the Waterloo WWTP were able to somewhat compensate for the effects of wastewater effluent. Prior to both plants' upgrades, the Waterloo and Kitchener WWTPs had similar ammonia concentrations in their effluent, but the Waterloo WWTP had much lower concentrations of many pharmaceuticals (Srikanthan 2019). Though there was not a strong correlation between effluent ammonia and sex steroid production (Appendix A4: Figure S1.14), perhaps the lower pharmaceutical concentrations could partially explain to the less-affected steroid response downstream of Waterloo. Although rainbow darter caught from DSW before the upgrades were completed had higher metabolic rates compared to those from REF2 (Mehdi et al. 2018), they may have still maintained enough energy to invest into androgen production. After the Waterloo WWTP upgrades rainbow darter had increased aerobic scope downstream of the effluent outfall (Hodgson 2020). It is possible that overall improvements in effluent quality (e.g., lower levels of ammonia and PPCPs in general) were responsible for the recovery in androgen production, rather than changes in estrogenicity, pharmaceuticals, or ammonia concentrations alone. However, the role of each of these stressors is difficult to separate in complex receiving environments.

Variability in *in vitro* steroid production was also observed among the three reference sites. These sites were chosen in order to assess background levels of steroid production, but they were not always consistent. For example, in 2012 11KT production at REF1 was very low, suggesting potential influence from the Shand Dam; however, this effect never repeated itself in subsequent years. The Shand Dam releases cold water which supports a tailwater brown trout fishery that extends as far downstream as REF1 (18 km away; Ontario Ministry of Natural Resources 2004). Therefore it is possible that in some years the cold water temperature delayed gonadal development and spawning, since rainbow darter need a temperature of at least 15°C to spawn (Reeves 1907). There was a significant positive correlation between temperature (May to October mean) and steroid production at REF1 (for both 11KT and T), although 2012 seemed to be anomalous year for low temperatures and low 11KT production (Appendix A4: Figure S1.15). However, this study did not specifically test this hypothesis and it is not known if a potential spawning delay in the spring or temperatures throughout the summer might affect steroid production in

the fall. It does, however, highlight how variability in biological endpoints can be greatly influenced by other factors and lead to changes that make linkages to specific stressors difficult to establish or interpret.

Steroid production was a reliable endpoint to use downstream of the Kitchener WWTP when the responses were strong and consistent (Marjan et al. 2018). However, it may not be sensitive enough when the response is weak, variable, and there are confounding effects upstream. Despite taking precautions to limit variability (e.g., sampling at the same time of year, ensuring there weren't confounding effects from handling or confinement stress) (Marjan et al. 2018), there was still considerable variability in the steroid responses around the Waterloo WWTP. However, it has been suggested that an increase in variability of an endpoint might also point to physiological disturbance, even if the mean or median value has not shifted (Hecker et al. 2002). Variability in steroid production in fish caught downstream of the WWTP could arise from fish seeking out refugia (Blanchfield et al. 2015). Even though rainbow darter generally have small home ranges (< 5 m; Hicks and Servos 2017), there is still a possibility that they could move out of the effluent plume, or that fish whose territories lie on the edge of the plume could be sometimes exposed to cleaner water depending on how the plume shifts over time. Overall, although it no longer seems to be impaired, steroid production was too variable to use for tracking the responses of rainbow darter to the Waterloo WWTP upgrades.

2.4.3 Intersex incidence and severity

There tended to be higher incidence and severity of intersex downstream of the Waterloo WWTP compared to the reference sites, especially during the WWTP upgrades while the effluent quality was poor. However, this was only statistically significant in some years as there was considerable variability at DSW. After the upgrades were completed, there were still some fish that showed intersex downstream of the Waterloo WWTP, but there were no significant differences in intersex incidence or severity score among sites. All sites (including the reference and downstream sites) had a consistent background level of intersex.

Prior to the upgrades, the maximum proportion of fish that were intersex downstream of the Waterloo WWTP was 55%. Fish from other wastewater-contaminated study systems show variable incidence of intersex depending on species and exposure (Bahamonde et al. 2013). For example, intersex in wild roach collected downstream of sewage effluents from various UK rivers ranged from 16% to 100% (Jobling et al. 1998, 2002), whereas intersex incidence in wild bream living downstream of multiple types of inputs (including wastewater) in the Elbe river in Germany was < 6% (Hecker et al. 2002). Within the Grand River, up to 100% of male fish were intersex downstream of the Kitchener

WWTP before the plant was upgraded (Hicks et al. 2017a). Despite differences in intersex severity, the recovery pattern was similar between the Kitchener and Waterloo WWTPs (Hicks et al. 2017a) where there was a decline in intersex in the years after the implementation of the aeration/nitrification upgrades.

Intersex incidence and severity in rainbow darter downstream of the Kitchener and Waterloo WWTPs after the upgrades at both plants were very similar (Hicks et al. 2017a). The timeline of the Waterloo recovery also followed the timeline reported after the Kitchener upgrades, with full recovery not observed until one to two seasons after the treatment upgrades (Hicks et al. 2017a). Rainbow darter begin to grow their gonads in the summer post-spawning, so if exposure occurs in this time period while they are sensitive, this may influence their gonadal development and expression of intersex even after exposure has been reduced (Hicks et al. 2017a).

Gonadal development of fish can be influenced by environmental factors such as water temperature, exogenous steroids, and other pollutants (Devlin and Nagahama 2002). Depending on the stage of development during which exposure takes place, intersex may be permanent or transitory (Devlin and Nagahama 2002). Studies suggest that fish are most sensitive to gonadal disruption during their early life stages but there is a great deal of variation in the type and severity of gonadal disruption observed which may depend on the concentration of EDCs, species, and exposure period (Devlin and Nagahama 2002). For example, a study that exposed fathead minnow to 10 ng/L EE2 during short windows in their development found ovarian-like ducts and cavity in male fish but no oocytes (van Aerle et al. 2002). A mesocosm experiment that exposed wild roach to wastewater effluent determined that early life (0 to 60 days post hatch) was the most sensitive life stage to influence from wastewater, and that intersex was a permanent condition in these fish even after depuration in clean water (Liney et al. 2005). The study also found that intersex was not induced in adult roach (Liney et al. 2005). However, intersex has been induced in adult male medaka after 6 weeks of exposure to a dose of 20 ng/L of EE2 (Hirakawa et al. 2012). Unpublished data suggest that rainbow darter will not develop intersex unless they are exposed in the early stages of their testes development (Hicks and Fuzzen, unpublished data). Therefore, the gonad development pattern of the species and the dose and timing of the exposure may be important factors in determining whether or not intersex can be induced. Chronic exposure to 5–6 ng/L of EE2 has been shown to result in intersex and the collapse of a wild fathead minnow population (Kidd et al. 2007). The population recovered within 3 years after the EE2 exposure was stopped, and testicular abnormalities were no longer present in the population (Blanchfield et al. 2015). Other species were not affected as severely in the study, and it was proposed that differences in life history (such as lifespan, preferred habitat, foraging location, and spawning times) can alter fishes' exposure and response to contaminants

(Palace et al. 2009). These factors may explain the differences in results often seen between the lab and the field and make it difficult to compare studies, especially when investigating different species with different patterns of gonad development. In a field study that collected wild fish, there was a significant increase in intersex wild roach caught downstream of sewage containing E2eq concentrations that were estimated to be above 1 ng/L (Jobling et al. 2006). These various results suggest species-specific or dose/environment-specific responses (Devlin and Nagahama 2002).

Previous modelling of the Waterloo WWTP effluent predicted that prior to the upgrades, under low flow conditions, fish may continue to be exposed to total estrogenicity (i.e., a benchmark value of 0.4 ng/L E2eq in the river) that would result in at least some intersex in fish downstream (Arlos et al. 2018). As estrogen loadings from the Waterloo WWTP declined after the upgrade, the impact on reproductive health of fish was expected to improve (Arlos et al. 2018; Srikanthan 2019). The current study confirmed these predictions. After the upgrades effluent E2eq concentrations dropped to below 1 ng/L (Srikanthan 2019). There was also a reduction but not a complete elimination of intersex in fish exposed to the effluent. These low levels of intersex were not surprising as there were background levels of intersex at the reference sites. However, the correlation between intersex and effluent estrogenicity in the current study was very weak (Appendix A5: Figure S1.17), suggesting that other components of the effluent (e.g., metformin (Niemuth and Klaper 2015) or other pharmaceuticals) might also have a role in inducing intersex in rainbow darter.

Years in which intersex incidence was particularly high or low may be partially explained by river flows and more generally by mean monthly/annual precipitation, but only anecdotally as there was no correlation between intersex incidence or severity and flow (Appendix A5: Figure S1.16). For example, in 2016 there were high levels of intersex and low river flows which may have exacerbated the effects of the wastewater effluent. Flows in 2013 were higher than average and there was low incidence of intersex despite poor-quality effluent. 2018 and 2019 were slightly drier than average but despite this there was still lower incidence of intersex at DSW, suggesting the positive influence of the improved effluent quality. Variable effluent quality likely confounded the potential relationship between river flow and intersex, but flow may have been important in altering exposure, especially during critical windows of sensitivity for intersex development in rainbow darter.

In general, intersex is a concern because it may have reproductive consequences for fish. Intersex has been shown to be related to reduced fertilization success in rainbow darter, but only for severely intersex fish (index 4–6) (Fuzzen et al. 2015). Despite this, survival to hatch was not found to be affected by intersex severity (Fuzzen et al. 2015). Additionally, Jobling et al. (2002) found that gonadal growth

was only inhibited in severely intersex wild roach. In the current study, severe intersex was rare, suggesting that in this reach of the Grand River, estrogen exposure associated with intersex is likely not inhibiting normal reproduction of rainbow darters.

Background levels of intersex of < 10% were observed at the reference sites (ranging from 0% to 20% each year). The reference sites used in this study were not pristine; the upper Grand River is heavily agricultural and also receives effluent from many smaller WWTPs (Loomer and Cooke 2011). Both of these inputs are a source of EDCs that may contribute to a low background level of intersex. Other study systems have also found intersex at reference sites. In England, one study reported intersex incidence in wild roach of 11.7% to 44.4% at upstream sites affected by dilute effluent and incidence of 4% to 18.1% at control sites (Jobling et al. 1998). This underscores the value of a long term study with multiple reference sites to be able to understand both the background variability and inter-annual variability before attempting to understand effects arising from a point source such as wastewater effluent (Bahamonde et al. 2013).

2.4.4 Conclusion

Biological responses in rainbow darter generally resembled reference site levels after the Waterloo WWTP was upgraded, but the some of the responses were highly variable and not always easily attributed directly to the wastewater effluent. The study system in the Grand River was complex with many types of inputs: agriculture, urban runoff, and wastewater effluent. Endpoints were also sometimes affected by environmental conditions including flow and water temperature, and these endpoints also likely respond on different exposure timelines, further complicating their relationship to the upgrades. Stable isotopes responded quickly to changes in effluent quality within months, whereas steroids and intersex took one to two years to respond. Endpoints that are on the lower end of the scale of biological organization tend to respond more quickly but may not be as ecologically relevant. Higher-level endpoints such as intersex take longer to respond, but since they are assimilated over a longer period of time they may better reflect exposure and the health of the individual and therefore be more suitable as a monitoring endpoint. However, endpoints at higher levels of biological organization are affected by many environmental conditions making it difficult to observe patterns or separate the influence of specific factors during monitoring (Fuzzen et al. 2016).

This study demonstrates the importance of long-term data collection, as anomalous weather or other unknown factors in some years may have led to misleading conclusions if not for a long-term baseline dataset. Additionally, multiple reference sites were important in order to have a more thorough

understanding of the baseline conditions, as well as to separate the influence of different sources of contaminants. Despite this, it may be prudent to use more sensitive endpoints, such as gene expression and metabolism, which may reflect changes in fish health more quickly. It is essential to choose appropriate and adequately sensitive endpoints in order to reduce the number of fish needed to detect differences and thus protect the natural population. It is also crucial to consider the ecological relevance of the endpoints, and their ability to inform researchers about a potential mechanistic or root cause of any disruption as well as to predict how a population may respond. Using an appropriate sample size of fish is important especially when the endpoints measured are highly variable or have a small effect size. The pre-upgrade conditions downstream of the Waterloo WWTP were less severe and more variable than the conditions downstream of the larger Kitchener WWTP, but similar trends were still apparent with improvements in effluent quality. Although the upgrades to these two plants targeted traditional effluent quality measures such as ammonia, there was also the co-benefit of a reduction in estrogenicity. These results demonstrate the environmental benefits that can be realized when process upgrades are implemented that generally improve effluent quality. The upgrades did not completely remove all contaminants, but the key endpoints of concern (e.g., *in vitro* steroid production, intersex) were reduced to levels that cannot be easily separated from upstream. Some effects may still be associated with the residual contaminant exposure in the effluent but overall, the investment in wastewater infrastructure has had a major effect on reducing effects in fish downstream.

Chapter 3
Conclusion and Recommendations

The aim of this thesis was to evaluate the biological responses of wild rainbow darter living upstream and downstream of the Waterloo WWTP to investigate if there were any changes associated with the WWTP upgrades. This study was motivated by the observation that rainbow darter showed recovery of several key biological responses after the upgrades to the Kitchener WWTP in 2012. In this study, stable isotope signatures were used to assess changes in effluent exposure in relation to the Waterloo WWTP upgrades. *In vitro* sex steroid production and intersex were measured to determine if infrastructure improvements to the Waterloo WWTP (improved aeration and nitrification of the effluent) contributed to a recovery of these endpoints.

The process upgrades at the Waterloo WWTP were implemented in order to improve traditional effluent quality parameters (i.e., ammonia) but were also effective at reducing the estrogenicity of the final effluent. The upgrades were associated with improvements in the health of wild rainbow darter living downstream as many endpoints are now indistinguishable from upstream reference conditions. However, it was often difficult to attribute these biological effects on fish directly to the changes in the WWTP effluent due to variability in the endpoints among the study sites, including the upstream reference sites. The Grand River is highly impacted by human activities (agriculture, urban development, dams, and wastewater) and there are many confounding factors present in the watershed including changes in river flow, inputs from agricultural runoff, other upstream WWTPs, and urban runoff, plus upsets to the treatment process during construction. Additionally, processes at the Waterloo WWTP are still being optimized and effluent quality monitoring continues to note improvements into 2020. Therefore, it remains possible that there could be a continued recovery in rainbow darter for the next few years, especially since new generations of fish will develop and reproduce in a less contaminated environment.

3.1 Efficacy of wastewater treatment upgrades

Standard aeration and nitrification upgrades aimed at improving effluent nutrient quality parameters seemed to be effective at improving the health of fish living downstream and this is good news for other municipalities that may also need to upgrade their WWTPs (e.g., to achieve secondary wastewater treatment by 2021 in Canada). Another factor that should be considered is the volume of effluent released

by the WWTP (i.e., the population serviced) relative to the base flow of the receiving river. It is important to consider fluctuations in flow throughout the year, especially low flow periods during the summer. This time period is critical since it is when fish may be exposed to the highest concentrations of effluent and may also be the most sensitive (although this may vary depending on the spawning time and mobility of the species present). Municipalities that treat larger volumes of wastewater may need to invest in higher levels of treatment in order to reduce contaminant loads. Additionally, the flow dynamics of the effluent plume, such as where it is located in the river, can affect the mixing of the effluent and subject some habitats to much higher concentrations of effluent than predicted based on the assumption of complete mixing.

3.2 Rainbow darter as a study species

Rainbow darter were an ideal sentinel species since they were abundant at all sampling locations and had a small home range. However, the endpoints that were measured were sometimes highly variable. For example, it was rare to find severe intersex in rainbow darters downstream of the Waterloo WWTP. This combined with the variable steroid data suggests that the wastewater effluent was arguably causing only a minor response in rainbow darter and not having an adverse or detrimental effect on their populations even before the upgrade. Rainbow darter are tolerant enough to be able to survive and reproduce in wastewater-contaminated areas and possibly even taking advantage of the minor changes in the environment such as greater nutrient availability. More sensitive species may have had a greater response to effluent exposure and then a more dramatic recovery following the upgrades. However, studies are limited by the abundance of fish species downstream of the WWTP, especially prior to the upgrades.

3.3 Environmental monitoring

There is currently no Environmental Effects Monitoring (EEM) program for wastewater effluents; however, rainbow darter in the Grand River are exposed to similar proportions of effluent that would trigger monitoring in other scenarios. EEM guidelines state that a fish survey is required if the concentration of metal mining or pulp and paper mill effluent is greater than 1% at a distance of 250 m downstream of the discharge point (Environment Canada 2010; 2012). If an EEM program for wastewater effluent were to be created, in addition to the standard endpoints of growth, body condition, LSI and GSI, it should also monitor estrogenic effects since estrogens have the potential to have an impact on the

sustainability of fish populations. Intersex may be an ideal endpoint to use in an EEM program since it has been demonstrated to be induced by exposure to low concentrations of estrogens (and other contaminants found in wastewater), is fairly sensitive, and has the potential to have direct implications for the health of the population.

3.4 Recommendation for future studies

This study was not initially designed as a monitoring project, but in hindsight it would have been ideal to have temperature loggers at each field site. Because many of the endpoints measured were reflective of a longer period of exposure (weeks to months), a single measurement taken at the time of fish sampling was not as useful compared to the temperature and flow logger data obtained from the GRCA and Water Survey of Canada, but unfortunately the logger data did not correspond perfectly to all of our sampling sites and could not be used in analyses at all of our sites. Additionally, having open communication with WWTP managers and operators was important for this study, and access to their weekly effluent nutrient dataset was extremely useful to support our analyses. Working in collaboration with wastewater managers and operators is critical for improved research outcomes.

A limitation of this study was the lack of ability to measure multiple endpoints within each individual fish. Rainbow darter gonads were often too small to split between analyses, so the steroid and intersex data are not always from the same fish and therefore could not be correlated. Performing correlations may have helped to better explore patterns between individual fish and may have explained some of the variability in the data. In terms of data analysis, conducting a before-after-control-impact (BACI) statistical analysis would have been ideal if there had been a more clearly defined before-after period. This was complicated in the current study because the WWTP underwent many changes over the years, as there were many construction delays and aeration tanks were commissioned at different times. Even after all the infrastructure upgrades went online, process optimization was still occurring until 2020. Because of this and considering the time it takes for rainbow darter to respond to changes in water quality, there were not yet enough post-upgrade years studied to date to do a BACI analysis.

Long-term studies are important and necessary to separate the effects of the stressor of interest from natural variability, especially when the stressor exerts subtle effects on the system. It was valuable to have multiple reference sites in order to characterize variability within the study system. Poor reference site selection would have led to different conclusions, and even the three reference sites chosen for this study did not always tell a consistent story. It was important to be aware of this and therefore not rely too

heavily on one reference site but instead consider overall patterns. It was also necessary to conduct field sampling at a consistent and correct time of year. Environmental changes throughout the year (such as seasonal flow and temperature patterns) affect the feasibility of sampling and also affect the responses of various endpoints. It is important to consider when endpoints (particularly reproductive endpoints such as steroid production) are at their highest and/or most stable level since there are natural variations over the course of a year related to spawning (Barrett and Munkittrick 2010; Tetreault et al. 2014). This is crucial for studies that take place over multiple years, because comparisons among years must be accurate. When studying multiple endpoints, there may not be a time when all of the endpoints are at their ideal time for sampling, so a compromise must be achieved. In this study, sampling occurred in the fall because it was a more stable period for gonadal development and because flows were too high in the spring, which would have led to inconsistencies in measurements of the endpoints.

As a follow up to the study performed, a laboratory experiment investigating the effects of different wastewater exposures would be helpful to further place these results into context and control for confounding environmental conditions (e.g., temperature, flow, dissolved oxygen levels, other sources of contamination, fish movement and avoidance, etc.). A caging study would also be useful to prevent fish from moving out of the effluent plume while retaining more natural environmental conditions. These kinds of additional studies would be particularly helpful for interpreting the *in vitro* steroid production data. Understanding more of the factors affecting rainbow darter steroid production and how these effects translate into higher levels of biological organization would be beneficial for putting these results into better ecological context. Gene expression of vitellogenin would be another valuable endpoint to measure since it is sensitive, has been linked to estrogen exposure, and was shown to be elevated in male rainbow darter prior to the Waterloo WWTP upgrades. I intended to include this endpoint and I was in the process of completing this work in the lab when COVID-19 resulted in the university lockdown.

3.5 Conclusions

Overall, this study adds to the body of literature investigating the long-term effects of WWTP effluent and the recovery of the receiving environment after facility upgrades. The findings from this work will support model development and validation and are important for supporting future improvements in water management policy and practice. These results may help inform best practices in wastewater management as more treatment plants require upgrades to meet regulations and/or ensure environmental protection. Improvements in water quality downstream of the Waterloo WWTP should also further reduce

contaminant exposure below the Kitchener outfall and should be investigated. Good wastewater treatment is increasingly important as the urban population grows, putting more pressure on ecosystems worldwide. This research supports the implementation of WWTP infrastructure upgrades in order to improve effluent quality and the health of fish living downstream.

References

Anderson C, Cabana G. 2006. Does δ15N in river food webs reflect the intensity and origin of N loads from the watershed? *Sci Total Environ* 367:968–978. https://doi.org/10.1016/j.scitotenv.2006.01.029

Angradi TR. 1994. Trophic linkages in the lower Colorado River: multiple stable isotope evidence. *J North Am Benthol Soc* 13:479–495. https://doi.org/10.2307/1467845

Arlos MJ, Parker WJ, Bicudo JR, et al. 2018. Multi-year prediction of estrogenicity in municipal wastewater effluents. *Sci Total Environ* 610–611:1103–1112. https://doi.org/10.1016/j.scitotenv.2017.08.171

Bahamonde PA, Fuzzen ML, Bennett CJ, et al. 2015a. Whole organism responses and intersex severity in rainbow darter (*Etheostoma caeruleum*) following exposures to municipal wastewater in the Grand River basin, ON, Canada. Part A. *Aquat Toxicol* 159:290–301. https://doi.org/10.1016/j.aquatox.2014.11.023

Bahamonde PA, McMaster ME, Servos MR, et al. 2015b. Molecular pathways associated with the intersex condition in rainbow darter (*Etheostoma caeruleum*) following exposures to municipal wastewater in the Grand River basin, ON, Canada. Part B. *Aquat Toxicol* 159:302–316. https://doi.org/10.1016/J.AQUATOX.2014.11.022

Bahamonde PA, Munkittrick KR, Martyniuk CJ. 2013. Intersex in teleost fish: Are we distinguishing endocrine disruption from natural phenomena? *Gen Comp Endocrinol* 192:25–35. https://doi.org/10.1016/j.ygcen.2013.04.005

Bahamonde PA, Tetreault GR, McMaster ME, et al. 2014. Molecular signatures in rainbow darter (*Etheostoma caeruleum*) inhabiting an urbanized river reach receiving wastewater effluents. *Aquat Toxicol* 148:211–220. https://doi.org/10.1016/J.AQUATOX.2014.01.010

Barber LB, Vajda AM, Douville C, et al. 2012. Fish endocrine disruption responses to a major wastewater treatment facility upgrade. *Environ Sci Technol* 46:2121–2131. https://doi.org/10.1021/es202880e

Barrett TJ, Munkittrick KR. 2010. Seasonal reproductive patterns and recommended sampling times for sentinel fish species used in environmental effects monitoring programs in Canada. *Environ Rev* 18:115–135. https://doi.org/10.1139/A10-004

Blanchfield PJ, Kidd KA, Docker MF, et al. 2015. Recovery of a wild fish population from whole-lake

additions of a synthetic estrogen. *Environ Sci Technol* 49:3136–3144. https://doi.org/10.1021/es5060513

Brown AR, Gunnarsson L, Kristiansson E, Tyler CR. 2014. Assessing variation in the potential susceptibility of fish to pharmaceuticals, considering evolutionary differences in their physiology and ecology. *Philos Trans R Soc B Biol Sci* 369:20130576–20130576. https://doi.org/10.1098/rstb.2013.0576

Brown CJM, Knight BW, McMaster ME, et al. 2011. The effects of tertiary treated municipal wastewater on fish communities of a small river tributary in Southern Ontario, Canada. *Environ Pollut* 159:1923–1931. https://doi.org/10.1016/j.envpol.2011.03.014

Cabana G, Rasmussen JB. 1996. Comparison of aquatic food chains using nitrogen isotopes. *Proc Natl Acad Sci U S A* 93:10844–10847

Cejudo E, Hood JL, Schiff SL, Aravena RO. 2018. Using the δ15N of submerged biomass for assessing changes in the nitrogen cycling in a river receiving wastewater treated effluent. *Ecol Indic* 95:645–653. https://doi.org/10.1016/j.ecolind.2018.08.013

Chambers PA, Allard M, Walker SL, et al. 1997. Impacts of municipal wastewater effluents on Canadian waters: A review. *Water Qual Res J Canada* 32:659–713. https://doi.org/10.2166/WQRJ.1997.038

Corcoran J, Winter MJ, Tyler CR. 2010. Pharmaceuticals in the aquatic environment: a critical review of the evidence for health effects in fish. *Crit Rev Toxicol* 40:287–304. https://doi.org/10.3109/10408440903373590

Crichton A. 2016. Age determination and growth of rainbow darter (Etheostoma caeruleum) in the Grand River, Ontario. MSc thesis. Department of Biology, University of Waterloo.

Daughton CG, Ternes TA. 1999. Pharmaceuticals and personal care products in the environment: Agents of subtle change? *Environ Health Perspect* 107:907–38. https://doi.org/10.1289/ehp.99107s6907

de Vlaming VL. 1972. Environmental control of teleost reproductive cycles: A brief review. *J Fish Biol* 4:131–140. https://doi.org/10.1111/j.1095-8649.1972.tb05661.x

deBruyn AMH, Rasmussen JB. 2002. Quantifying assimilation of sewage-derived organic matter by riverine benthos. *Ecol Appl* 12:511–520. https://doi.org/10.1890/1051-0761(2002)012[0511:QAOSDO]2.0.CO;2

Devlin RH, Nagahama Y. 2002. Sex determination and sex differentiation in fish: an overview of genetic,

physiological, and environmental influences. *Aquaculture* 208:191–364. https://doi.org/10.1016/S0044-8486(02)00057-1

Environment Canada. 2001. The state of municipal wastewater effluents in Canada. Environment Canada, Ottawa, Ontario

Environment Canada. 2012. Metal Mining Technical Guidance for Environmental Effects Monitoring. https://www.canada.ca/en/environment-climate-change/services/managing-pollution/environmental-effects-monitoring/pulp-paper-technical-guidance.html

Environment Canada. 2010. 2010 Pulp and Paper Environmental Effects Monitoring (EEM) Technical Guidance Document. https://www.canada.ca/content/dam/eccc/migration/main/esee-eem/3e389bd4-e48e-4301-a740-171c7a887ee9/pp_full_versionenglish-5B1-5D-final-2.0.pdf

Finlay JC. 2004. Patterns and controls of lotic algal stable carbon isotope ratios. *Limnol Oceanogr* 49:850–861. https://doi.org/10.4319/lo.2004.49.3.0850

Finlay JC, Power ME, Cabana G. 1999. Effects of water velocity on algal carbon isotope ratios: implications for river food web. *Limnol Oceanogr* 44:1198–1203

Folmar LC, Denslow ND, Rao V, et al. 1996. Vitellogenin induction and reduced serum testosterone concentrations in feral male carp (*Cyprinus carpio*) captured near a major metropolitan sewage treatment plant. *Environ Health Perspect* 104:1096. https://doi.org/10.2307/3433123

Fox J, Weisberg S. 2019. An R Companion to Applied Regression, Third Edition. Sage: Thousand Oaks, CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

France RL. 1995a. Critical examination of stable isotope analysis as a means for tracing carbon pathways in stream ecosystems. *Can J Fish Aquat Sci* 52:651–656. https://doi.org/10.1139/f95-065

France RL. 1995b. Carbon-13 enrichment in benthic compared to planktonic algae: foodweb implications. *Mar Ecol Prog Ser* 124:307–312. https://doi.org/doi:10.3354/meps124307

Fuzzen MLM, Bennett CJ, Tetreault GR, et al. 2015. Severe intersex is predictive of poor fertilization success in populations of rainbow darter (*Etheostoma caeruleum*). *Aquat Toxicol* 160:106–116. https://doi.org/10.1016/J.AQUATOX.2015.01.009

Fuzzen MLM, Bragg LM, Tetreault GR, et al. 2016. An assessment of the spatial and temporal variability of biological responses to municipal wastewater effluent in rainbow darter (*Etheostoma caeruleum*) collected along an urban gradient. *PLoS One* 11:e0164879.

https://doi.org/10.1371/journal.pone.0164879

Galloway BJ, Munkittrick KR. 2006. Influence of seasonal changes in relative liver size, condition, relative gonad size and variability in ovarian development in multiple spawning fish species used in environmental monitoring programmes. *J Fish Biol* 69:1788–1806. https://doi.org/10.1111/j.1095-8649.2006.01249.x

Gillis PL. 2012. Cumulative impacts of urban runoff and municipal wastewater effluents on wild freshwater mussels (*Lasmigona costata*). *Sci Total Environ* 431:348–356. https://doi.org/10.1016/j.scitotenv.2012.05.061

Grand River Conservation Authority. 2014. Grand River Watershed Water Management Plan. Cambridge, ON

Grand River Conservation Authority. 2020. Monitoring Downloads - GRIN. In: Monit. Data Downloads. https://data.grandriver.ca/downloads-monitoring.html. Accessed 24 Apr 2020

Gunnarsson L, Jauhiainen A, Kristiansson E, et al. 2008. Evolutionary conservation of human drug targets in organisms used for environmental risk assessments. *Environ Sci Technol* 42:5807–5813. https://doi.org/10.1021/es8005173

Harris CA, Hamilton PB, Runnalls TJ, et al. 2011. The consequences of feminization in breeding groups of wild fish. *Environ Health Perspect* 119:306–311. https://doi.org/10.1289/ehp.1002555

Heaton THE. 1986. Isotopic studies of nitrogen pollution in the hydrosphere and atmosphere: a review. *Chem Geol (Isotope Geosci Sect)* 59:87–102. https://doi.org/10.1016/0168-9622(86)90059-X

Hecker M, Tyler CR, Hoffmann M, et al. 2002. Plasma biomarkers in fish provide evidence for endocrine modulation in the Elbe river, Germany. *Environ Sci Technol* 36:2311–2321. https://doi.org/10.1021/es010186h

Hicks KA. 2017. Response of wild fish to municipal wastewater treatment plant upgrades. PhD thesis. Department of Biology, University of Waterloo. http://hdl.handle.net/10012/11569

Hicks KA, Fuzzen MLM, McCann EK, et al. 2017a. Reduction of intersex in a wild fish population in response to major municipal wastewater treatment plant upgrades. *Environ Sci Technol* 51:1811–1819. https://doi.org/10.1021/acs.est.6b05370

Hicks KA, Loomer HA, Fuzzen MLM, et al. 2017b. δ15N tracks changes in the assimilation of sewage-derived nutrients into a riverine food web before and after major process alterations at two

municipal wastewater treatment plants. *Ecol Indic* 72:747–758. https://doi.org/10.1016/j.ecolind.2016.09.011

Hicks KA, Servos MR. 2017. Site fidelity and movement of a small-bodied fish species, the rainbow darter (*Etheostoma caeruleum*): Implications for environmental effects assessment. *River Res Appl* 33:1016–1025. https://doi.org/10.1002/rra.3161

Hirakawa I, Miyagawa S, Katsu Y, et al. 2012. Gene expression profiles in the testis associated with testis-ova in adult Japanese medaka (*Oryzias latipes*) exposed to 17α-ethinylestradiol. *Chemosphere* 87:668–674. https://doi.org/10.1016/j.chemosphere.2011.12.047

Hodgson RM. 2020. Changes in gill physiology and energy requirements of darter species (*Etheostoma* spp.) due to effluent in the Grand River. MSc thesis. Department of Biology, University of Waterloo.

Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70. https://doi.org/10.2307/4615733

Hughes BB, Beas-Luna R, Barner AK, et al. 2017. Long-term studies contribute disproportionately to ecology and policy. *Bioscience* 67:271–278. https://doi.org/10.1093/biosci/biw185

Jardine JJ, Van Der Kraak GJ, Munkittrick KR. 1996. Capture and confinement stress in white sucker exposed to bleached kraft pulp mill effluent. *Ecotoxicol Environ Saf* 33:287–298. https://doi.org/10.1006/eesa.1996.0037

Jardine TD, Kidd KA, Fisk AT. 2006. Applications, considerations, and sources of uncertainty when using stable isotope analysis in ecotoxicology. *Environ Sci Technol* 40:7501–7511. https://doi.org/10.1021/ES061263H

Jobling S, Beresford N, Nolan M, et al. 2002. Altered sexual maturation and gamete production in wild roach (*Rutilus rutilus*) living in rivers that receive treated sewage effluents. *Biol Reprod* 66:272–281. https://doi.org/10.1095/biolreprod66.2.272

Jobling S, Nolan M, Tyler CR, et al. 1998. Widespread sexual disruption in wild fish. *Environ Sci Technol* 32:2498–2506. https://doi.org/10.1021/es9710870

Jobling S, Williams R, Johnson A, et al. 2006. Predicted exposures to steroid estrogens in U.K. rivers correlate with widespread sexual disruption in wild fish populations. *Environ Health Perspect* 114:32–39. https://doi.org/10.1289/ehp.8050

Kaushal SS, Likens GE, Jaworski NA, et al. 2010. Rising stream and river temperatures in the United States. *Ecol Soc Am* 8:461–466. https://doi.org/10.1890/090037

Kidd KA, Blanchfield PJ, Mills KH, et al. 2007. Collapse of a fish population after exposure to a synthetic estrogen. *Proc Natl Acad Sci* 104:8897–8901. https://doi.org/10.1073/pnas.0609568104

Kidd KA, Paterson MJ, Rennie MD, et al. 2014. Direct and indirect responses of a freshwater food web to a potent synthetic oestrogen. *Philos Trans R Soc B Biol Sci* 369:20130578. https://doi.org/10.1098/rstb.2013.0578

Liney KE, Hagger JA, Tyler CR, et al. 2006. Health effects in fish of long-term exposure to effluents from wastewater treatment works. *Environ Health Perspect* 114:81–89. https://doi.org/10.1289/ehp.8058

Liney KE, Jobling S, Shears JA, et al. 2005. Assessing the sensitivity of different life stages for sexual disruption in roach (*Rutilus rutilus*) exposed to effluents from wastewater treatment works. *Environ Health Perspect* 113:1299–1307. https://doi.org/10.1289/ehp.7921

Loomer HA. 2008. The dynamics of carbon and nitrogen stable isotope analysis of aquatic organisms within the Grand River Watershed. MSc thesis. Department of Biology, University of Waterloo. http://hdl.handle.net/10012/4263

Loomer HA, Cooke SE. 2011. Water quality in the Grand River watershed: current conditions & trends (2003-2008). Gd River Conserv Authority, Cambridge, Ontario

Loomer HA, Oakes KD, Schiff SL, et al. 2015. Use of stable isotopes to trace municipal wastewater effluents into food webs within a highly developed river system. *River Res Appl* 31:1093–1100. https://doi.org/10.1002/rra.2826

Luo Y, Guo W, Ngo HH, et al. 2014. A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. *Sci Total Environ* 473–474:619–641. https://doi.org/10.1016/j.scitotenv.2013.12.065

Mangiafico S. 2020. rcompanion: Functions to Support Extension Education Program Evaluation. R package version 2.3.25. https://cran.r-project.org/package=rcompanion

Marjan P, Martyniuk CJ, Fuzzen MLM, et al. 2017. Returning to normal? Assessing transcriptome recovery over time in male rainbow darter (*Etheostoma caeruleum*) liver in response to wastewater-treatment plant upgrades. *Environ Toxicol Chem* 36:2108–2122. https://doi.org/10.1002/etc.3741

Marjan P, Van Der Kraak GJ, MacLatchy DL, et al. 2018. Assessing recovery of in vitro steroid production in male rainbow darter (*Etheostoma caeruleum*) in response to municipal wastewater treatment plant infrastructure changes. *Environ Toxicol Chem* 37:501–514. https://doi.org/10.1002/etc.3986

Martinović D, Hogarth WT, Jones RE, Sorensen PW. 2007. Environmental estrogens suppress hormones, behavior, and reproductive fitness in male fathead minnows. *Environ Toxicol Chem* 26:271–8

Matthiessen P, Wheeler JR, Weltje L. 2018. A review of the evidence for endocrine disrupting effects of current-use chemicals on wildlife populations. *Crit Rev Toxicol* 48:195–216. https://doi.org/10.1080/10408444.2017.1397099

McCallum ES, Nikel KE, Mehdi H, et al. 2019. Municipal wastewater effluent affects fish communities: A multi-year study involving two wastewater treatment plants. *Environ Pollut* 252:. https://doi.org/10.1016/j.envpol.2019.06.075

McMaster ME, Munkittrick KR, Jardine JJ, et al. 1995. Protocol for measuring *in vitro* steroid production by fish gonadal tissue

Mehdi H, Dickson FH, Bragg LM, et al. 2018. Impacts of wastewater treatment plant effluent on energetics and stress response of rainbow darter (*Etheostoma caeruleum*) in the Grand River watershed. *Comp Biochem Physiol Part - B Biochem Mol Biol* 224:270–279. https://doi.org/10.1016/j.cbpb.2017.11.011

Mills LJ, Chichester C. 2005. Review of evidence: Are endocrine-disrupting chemicals in the aquatic environment impacting fish populations? *Sci Total Environ* 343:1–34. https://doi.org/10.1016/j.scitotenv.2004.12.070

Ministry of Justice. 2012. Wastewater systems effluent regulations (SOR SOR/2012-139). In: Fish. Act. https://laws-lois.justice.gc.ca/eng/regulations/sor-2012-139/index.html. Accessed 22 Apr 2020

Morrissey CA, Boldt A, Mapstone A, et al. 2013. Stable isotopes as indicators of wastewater effects on the macroinvertebrates of urban rivers. *Hydrobiologia* 700:231–244. https://doi.org/10.1007/s10750-012-1233-7

Munkittrick KR, Arens CJ, Lowell RB, Kaminski GP. 2009. A review of potential methods of determining critical effect size for designing environmental monitoring programs. *Environ Toxicol Chem* 28:1361–1371. https://doi.org/10.1897/08-376.1

Nelson KC, Palmer MA. 2007. Stream temperature surges under urbanization and climate change: data, models, and responses. *J Am Water Resour Assoc* 43:440–452. https://doi.org/10.1111

Niemuth NJ, Klaper RD. 2015. Emerging wastewater contaminant metformin causes intersex and reduced fecundity in fish. *Chemosphere* 135:38–45. https://doi.org/10.1016/j.chemosphere.2015.03.060

Odjadjare EEO, Okoh AI. 2010. Physicochemical quality of an urban municipal wastewater effluent and its impact on the receiving environment. *Environ Monit Assess* 170:383–394. https://doi.org/10.1007/s10661-009-1240-y

Ontario Ministry of Natural Resources. 2004. Grand River Tailwater Fisheries Tailwater Fisheries Management Plan: 2005-2010. https://www.grandriver.ca/en/our-watershed/resources/Documents/Fishery/Fishery_TailwaterPlan.pdf. Accessed 17 Apr 2020

Palace VP, Evans RE, Wautier KG, et al. 2009. Interspecies differences in biochemical, histopathological, and population responses in four wild fish species exposed to ethynylestradiol added to a whole lake. *Can J Fish Aquat Sci* 66:1920–1935. https://doi.org/10.1139/F09-125

Parrott JL, Blunt BR. 2005. Life-cycle exposure of fathead minnows (*Pimephales promelas*) to an ethinylestradiol concentration below 1 ng/L reduces egg fertilization success and demasculinizes males. *Environ Toxicol* 20:131–141. https://doi.org/10.1002/tox.20087

Peterson BJ, Fry B. 1987. Stable isotopes in ecosystem studies. *Annu Rev Ecol Syst* 18:293–320. https://doi.org/10.1146/annurev.es.18.110187.001453

Petrie B, Barden R, Kasprzyk-Hordern B. 2015. A review on emerging contaminants in wastewaters and the environment: Current knowledge, understudied areas and recommendations for future monitoring. *Water Res* 72:3–27. https://doi.org/10.1016/j.watres.2014.08.053

Pitt KA, Connolly RM, Maxwell P. 2009. Redistribution of sewage-nitrogen in estuarine food webs following sewage treatment upgrades. *Mar Pollut Bull* 58:573–580. https://doi.org/10.1016/j.marpolbul.2008.11.016

Post DM. 2002. Using stable isotopes to estimate trophic position: Models, methods, and assumptions. *Ecology* 83:703–718. https://doi.org/10.1890/0012-9658(2002)083[0703:USITET]2.0.CO;2

R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Rasmussen JB, Trudeau V. 2010. How well are velocity effects on δ13C signatures transmitted up the

food web from algae to fish? *Freshw Biol* 55:1303–1314. https://doi.org/10.1111/j.1365-2427.2009.02354.x

Rasmussen JB, Trudeau V. 2007. Influence of velocity and chlorophyll standing stock on periphyton δ13C and δ15N in the Ste. Marguerite River system, Quebec. *Can J Fish Aquat Sci* 64:1370–1381. https://doi.org/10.1139/F07-109

Reeves CD. 1907. The breeding habits of the rainbow darter (*Etheostoma cæruleum* Storer), a study in sexual selection. *Biol Bull* 14:35–59. https://doi.org/10.2307/1535646

Region of Waterloo. 2018. Region of Waterloo 2018 Wastewater Treatment Master Plan. Kitchener, Ontario. https://www.regionofwaterloo.ca/en/living-here/resources/Documents/water/projects/wastewater/plan/WS2018Wastewater_Treatment_Master_Plan_WWTMP_Final_Report.pdf

Region of Waterloo. 2019. Water and Wastewater Monitoring Report 2019. Kitchener, Ontario. https://www.regionofwaterloo.ca/en/living-here/resources/Documents/water/reports/WS2019WaterAndWastewaterMonitoringReport.pdf

Reid AJ, Carlson AK, Creed IF, et al. 2018. Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biol Rev* 94:849–873. https://doi.org/10.1111/brv.12480

Robinson CS, Tetreault GR, McMaster ME, Servos MR. 2016. Impacts of a tertiary treated municipal wastewater effluent on the carbon and nitrogen stable isotope signatures of two darter species (*Etheostoma blennioides* and *E. caeruleum*) in a small receiving environment. *Ecol Indic* 60:594–602. https://doi.org/10.1016/j.ecolind.2015.06.041

Rojas MR, Leung C, Bonk F, et al. 2013. Assessment of the effectiveness of secondary wastewater treatment technologies to remove trace chemicals of emerging concern. *Crit Rev Environ Sci Technol* 43:1281–1314. https://doi.org/10.1080/10643389.2011.644221

Salierno JD, Kane AS. 2009. 17α-ethinylestradiol alters reproductive behaviors, circulating hormones, and sexual morphology in male fathead minnows *(Pimephales promelas)*. *Environ Toxicol Chem* 28:953. https://doi.org/10.1897/08-111.1

Servos MR, Bennie DT, Burnison BK, et al. 2005. Distribution of estrogens, 17β-estradiol and estrone, in Canadian municipal wastewater treatment plants. *Sci Total Environ* 336:155–170. https://doi.org/10.1016/J.SCITOTENV.2004.05.025

Söffker M, Tyler CR. 2012. Endocrine disrupting chemicals and sexual behaviors in fish – a critical review on effects and possible consequences. *Crit Rev Toxicol* 42:653–668. https://doi.org/10.3109/10408444.2012.692114

Srikanthan N. 2019. Analysis of temporal changes in estrogenic compounds released from municipal wastewater treatment plants. MSc thesis. Department of Biology, University of Waterloo. http://hdl.handle.net/10012/15167. University of Waterloo

Steffy LY, Kilham SS. 2004. Elevated δ15N in stream biota in areas with septic tank systems in an urban watershed. *Ecol Appl* 14:637–641. https://doi.org/10.1890/03-5148

Tanna RN, Tetreault GR, Bennett CJ, et al. 2013. Occurrence and degree of intersex (testis-ova) in darters (*Etheostoma* spp.) across an urban gradient in the Grand River, Ontario, Canada. *Environ Toxicol Chem* 32:1981–1991. https://doi.org/10.1002/etc.2262

Ternes TA, Stumpf M, Mueller J, et al. 1999. Behavior and occurrence of estrogens in municipal sewage treatment plants - I. Investigations in Germany, Canada and Brazil. *Sci Total Environ* 225:81–90

Tetreault GR, Bennett CJ, Servos MR, McMaster ME. 2014. Optimization of effects-assessment of greenside darter (*Etheostoma blennioides*) exposed to tertiary treated municipal wastewater based on seasonal changes of reproductive endpoints. *Environ Toxicol Chem* 33:1077–1089. https://doi.org/10.1002/etc.2526

Tetreault GR, Bennett CJ, Shires K, et al. 2011. Intersex and reproductive impairment of wild fish exposed to multiple municipal wastewater discharges. *Aquat Toxicol* 104:278–290. https://doi.org/10.1016/J.AQUATOX.2011.05.008

Tetreault GR, Jm Brown C, Bennett J, et al. 2013. Fish community responses to multiple municipal wastewater inputs in a watershed. *Integr Environ Assess Manag* 9:456–468. https://doi.org/10.1002/ieam.1364

Tyler CR, Jobling S. 2008. Roach, sex, and gender-bending chemicals: the feminization of wild fish in English rivers. *Bioscience* 58:1051–1059. https://doi.org/10.1641/B581108

Tyler CR, Jobling S, Sumpter JP. 1998. Endocrine disruption in wildlife: a critical review of the evidence. *Crit Rev Toxicol* 28:319–361. https://doi.org/10.1080/10408449891344236

Ulseth AJ, Hershey AE. 2005. Natural abundances of stable isotopes trace anthropogenic N and C in an urban stream. *J North Am Benthol Soc* 24:270–289. https://doi.org/10.1899/03-080.1

Vajda AM, Barber LB, Gray JL, et al. 2008. Reproductive disruption in fish downstream from an estrogenic wastewater effluent. *Environ Sci Technol* 42:3407–3414. https://doi.org/10.1021/es0720661

Vajda AM, Barber LB, Gray JL, et al. 2011. Demasculinization of male fish by wastewater treatment plant effluent. *Aquat Toxicol* 103:213–221. https://doi.org/10.1016/J.AQUATOX.2011.02.007

van Aerle R, Pounds N, Hutchinson TH, et al. 2002. Window of sensitivity for the estrogenic effects of ethinylestradiol in early life-stages of fathead minnow, *Pimephales promelas*. *Ecotoxicology* 11:423–434. https://doi.org/10.1023/A:1021053217513

Wang J, Wang S. 2016. Removal of pharmaceuticals and personal care products (PPCPs) from wastewater: a review. *J Environ Manage* 182:620–640. https://doi.org/10.1016/j.jenvman.2016.07.049

Water Survey of Canada. 2020. Historical Hydrometric Data Search. In: Natl. Hydrol. Serv. Environ. Clim. Chang. Canada. https://wateroffice.ec.gc.ca/mainmenu/historical_data_index_e.html

Weber AA, Moreira DP, Melo RMC, et al. 2019. Environmental exposure to oestrogenic endocrine disruptors mixtures reflecting on gonadal sex steroids and gametogenesis of the neotropical fish *Astyanax rivularis*. *Gen Comp Endocrinol* 279:99–108. https://doi.org/10.1016/j.ygcen.2018.12.016

Wickham H, Averick M, Bryan J, et al. 2019. Welcome to the tidyverse. *J Open Source Softw* 4:1686. https://doi.org/10.21105/joss.01686

Winn HE. 1958. Comparative reproductive behavior and ecology of fourteen species of darters (Pisces-Percidae). *Ecol Monogr* 28:155–191. https://doi.org/10.2307/1942207

Appendix A

Supplemental Information

A1 Water and effluent quality

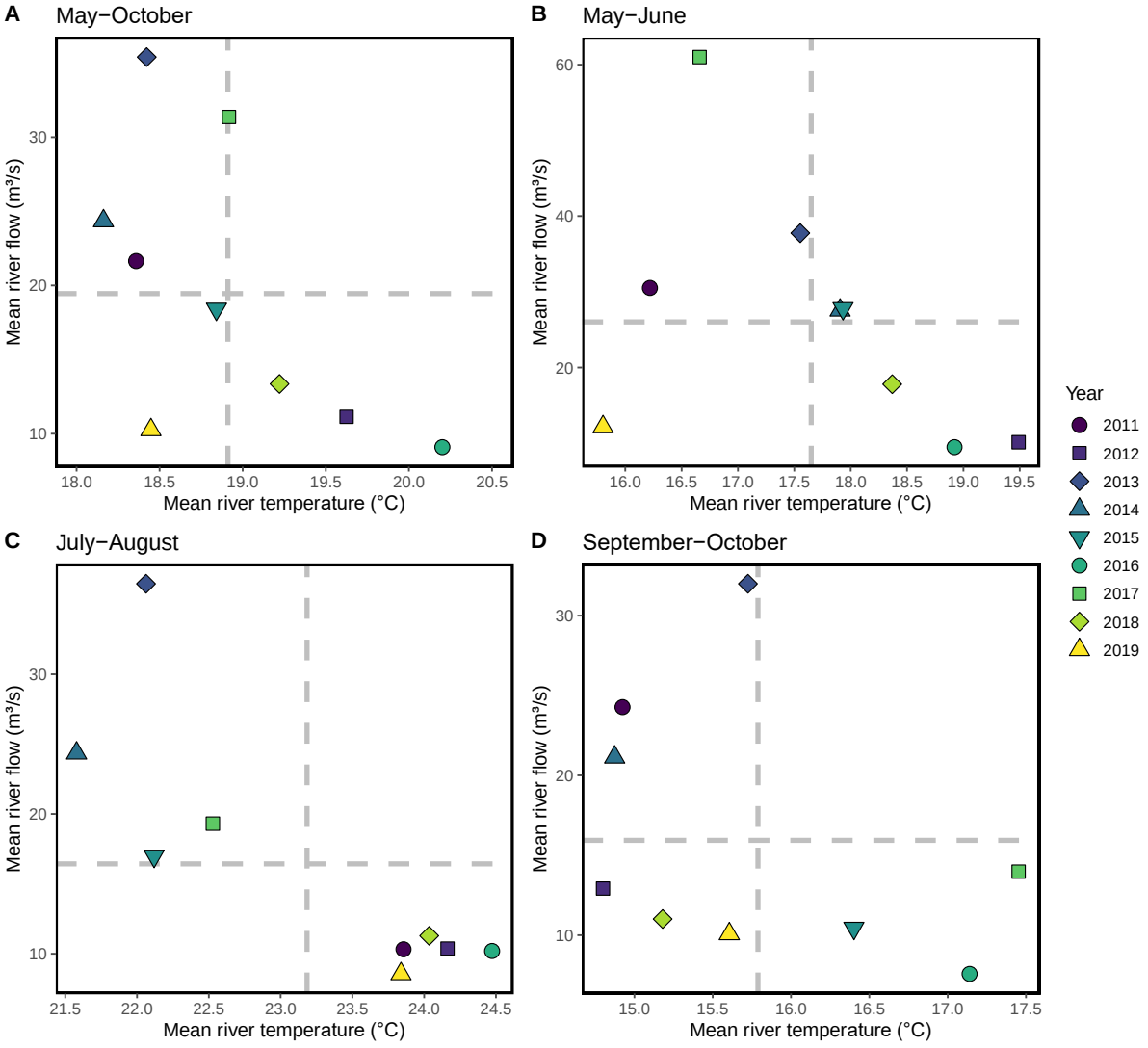

Figure S1.1. River temperature and flow at Bridgeport monitoring station (GRCA station 68; located 0.85 km upstream of the Waterloo WWTP outfall). (A) May to October averages. (B) Spring (May and June) averages. (C) Summer (July and August) averages. (D) Fall (September and October) averages. For each panel, the dashed grey lines indicate the mean flow and temperature for all years.

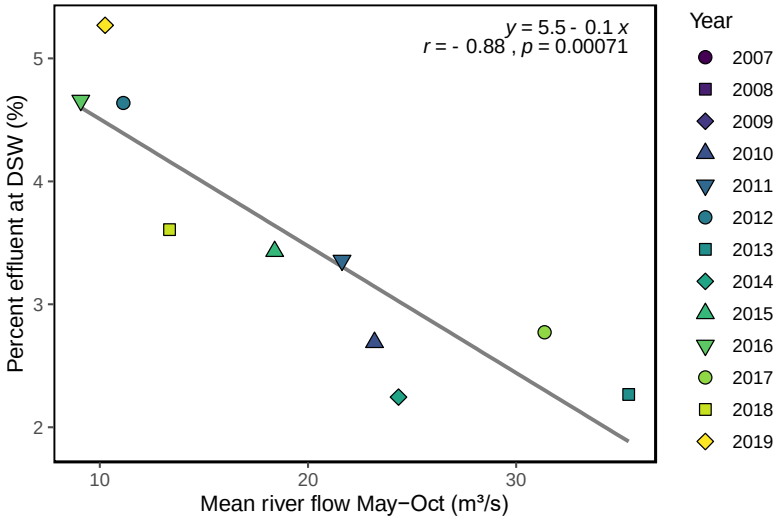

Figure S1.2 Correlation between mean percent effluent exposure at DSW and mean river flow at the Bridgeport monitoring station (GRCA number 68), located 0.85 km upstream of the WWTP outfall (Pearson's $r = -0.88$, $p < 0.001$). Mean river flow was calculated from the average of the daily mean flow from May to October of each year. Percent effluent was calculated assuming instantaneous mixing of effluent and river water, though it is known that mixing is not fully complete by the time the water reaches DSW and therefore fish may in reality be exposed to greater concentrations of effluent than indicated.

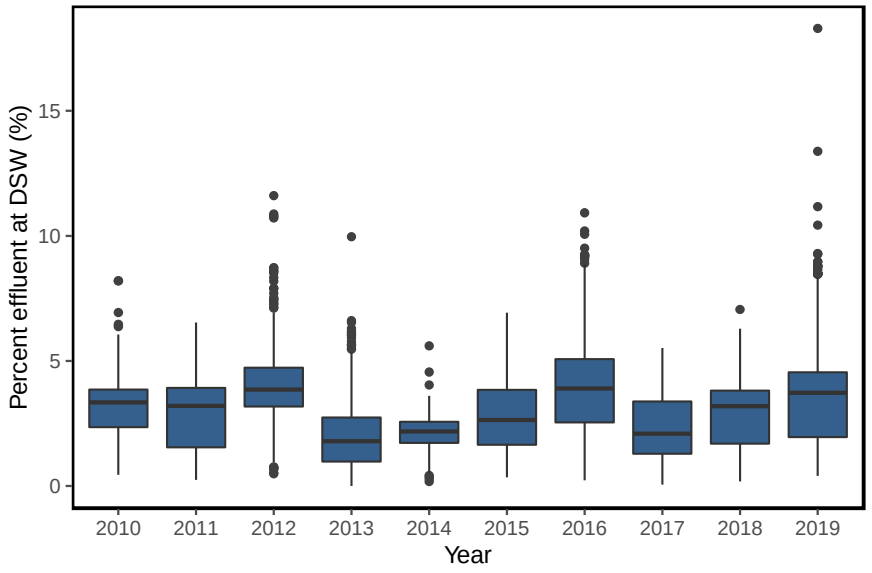

Figure S1.3 Boxplots showing annual estimated percent effluent exposure downstream of the Waterloo WWTP, calculated using daily measurements of effluent and river flow. Percent effluent exposure was calculating assuming an instantaneous mixing of effluent and river water, so the actual amount of effluent that fish are exposed to is likely higher than the graph indicates.

A2 Conestogo River sites

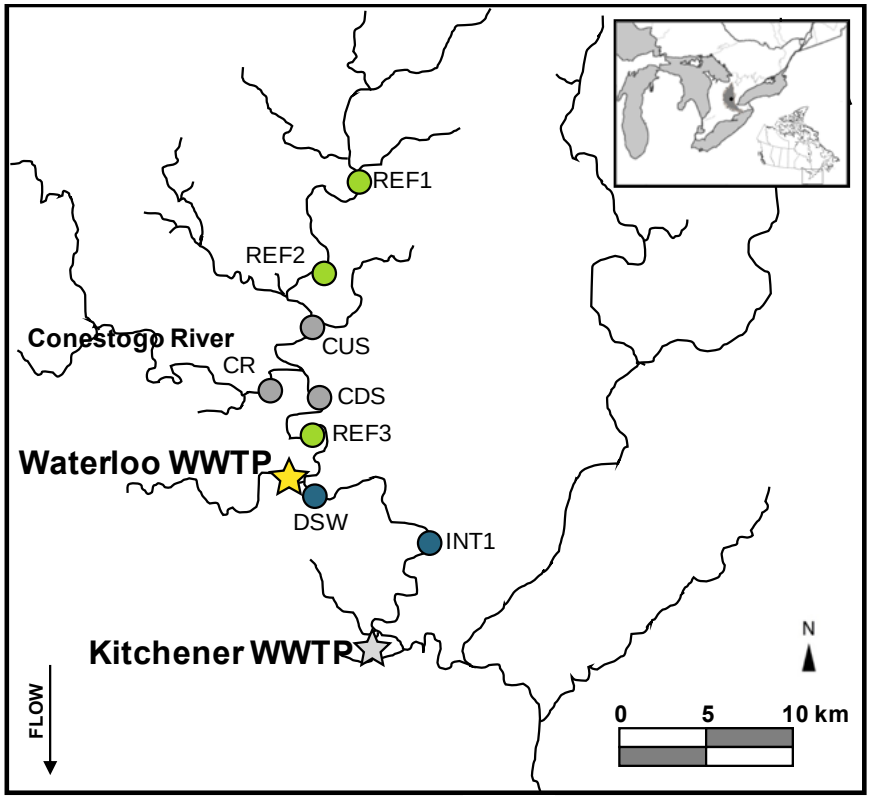

Figure S1.4 Map of the Grand River with sites sampled for the main study as well as the extra sites sampled along the Conestogo River in 2015 and 2018. CUS is in the Grand River upstream of the Conestogo River confluence, CR is located in the Conestogo River, and CDS is in the Grand River downstream of the Conestogo River confluence.

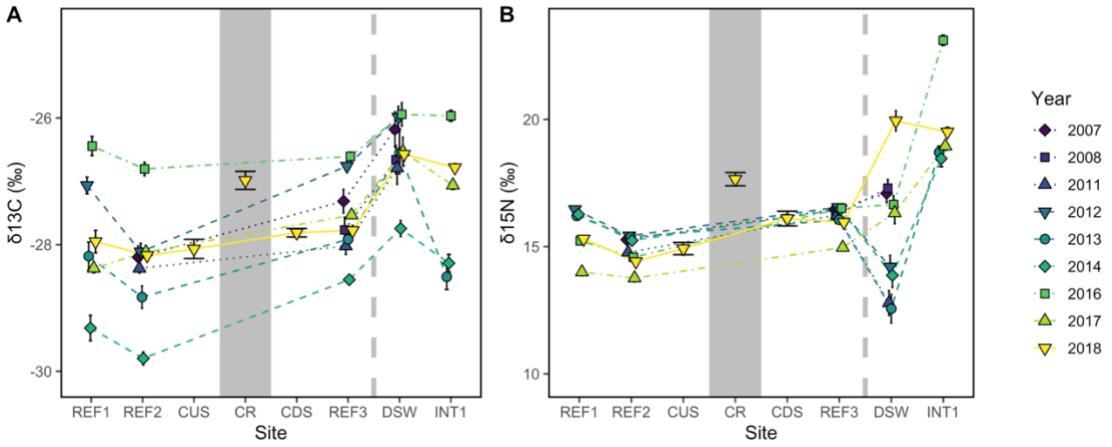

Figure S1.5 (A) $\delta^{13}C$ and (B) $\delta^{15}N$ along the Grand River, including the extra Conestogo River sites sampled in 2018. CUS is in the Grand River upstream of the Conestogo River confluence, CR is located in the Conestogo River, and CDS is in the Grand River downstream of the Conestogo River confluence. The shaded region of both graphs indicates the site sampled along the Conestogo River, whereas the white background is the sites sampled along the Grand River. $\delta^{13}C$ and $\delta^{15}N$ are both more enriched in the Conestogo River but this effect is diluted by the confluence.

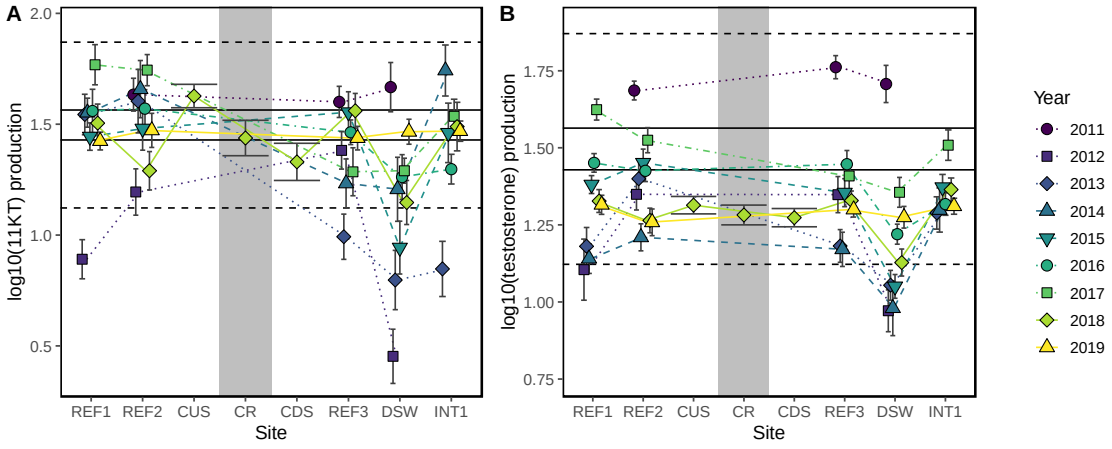

Figure S1.6 *In vitro* production of (A) 11KT and (B) testosterone by rainbow darter caught at sites along the Grand River, including the additional sites added around the Conestogo River in 2018 CUS is in the Grand River upstream of the Conestogo River confluence, CR is located in the Conestogo River, and CDS is in the Grand River downstream of the Conestogo River confluence. The shaded region of both graphs indicates the site sampled along the Conestogo River, whereas the white background is the sites sampled along the Grand River. There were no distinct patterns in steroid production among sites in 2018.

A3 Stable isotopes

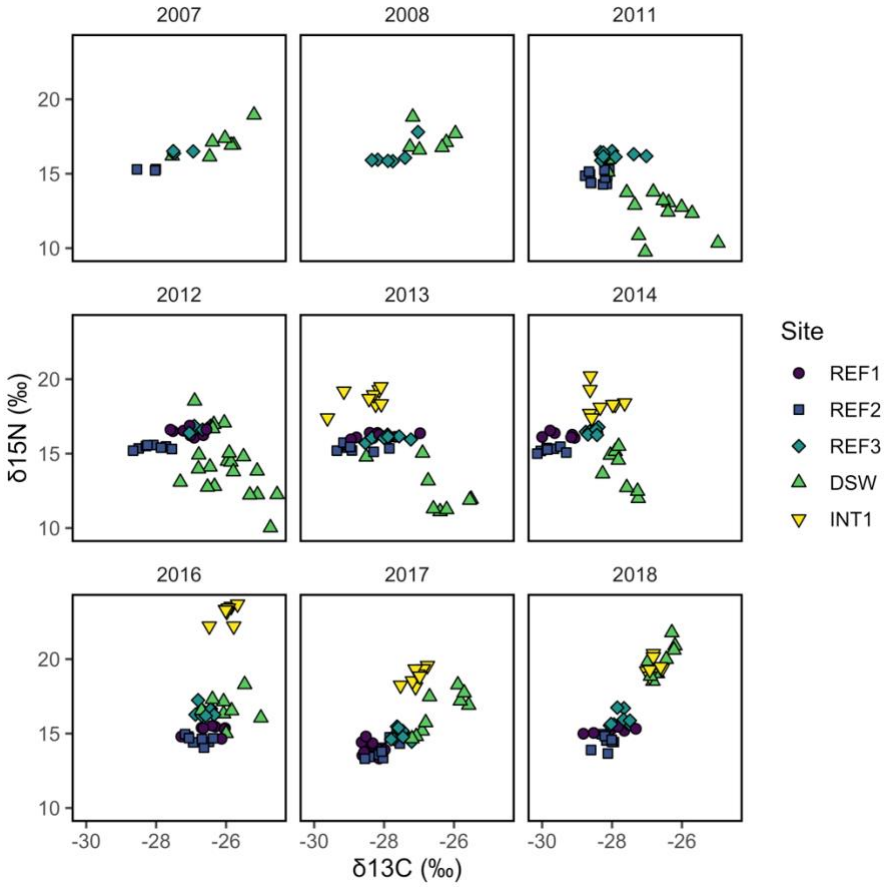

Figure S1.7 Biplots of $\delta^{15}N$ and $\delta^{13}C$ signatures in rainbow darter at all sites each year. During the period of poor-quality effluent during the upgrades (2011–2014), DSW had lower $\delta^{15}N$ compared to the reference sites.

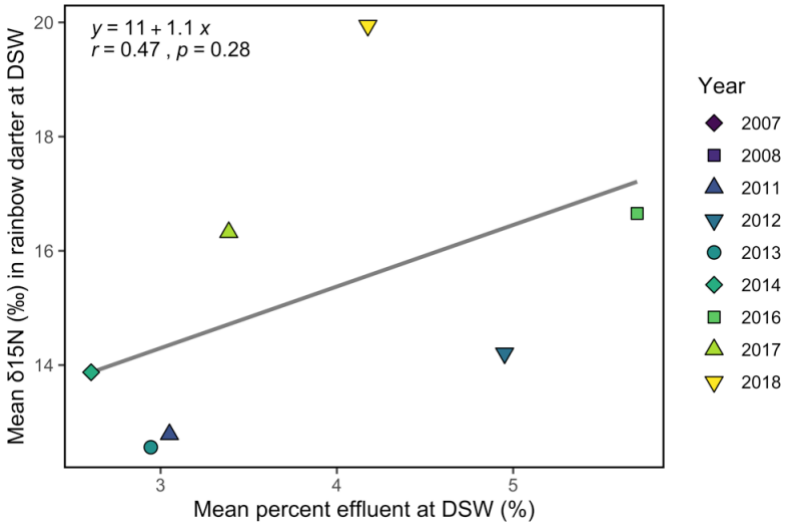

Figure S1.8 Correlation between the mean $\delta^{15}N$ at DSW and the mean percent effluent exposure from September to October each year. Percent effluent exposure was calculating assuming instantaneous mixing of effluent and river water, so the actual amount of effluent that fish are exposed to is likely higher than is indicated by the graph.

A4 *In vitro* steroid production

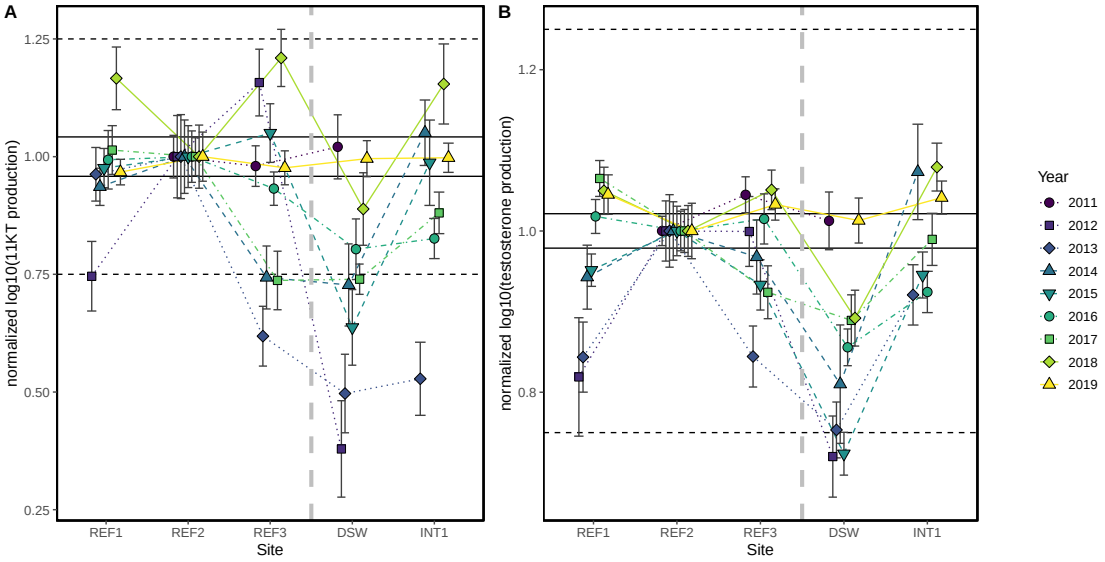

Figure S1.9 Normalized logged steroid production (mean ± SE) of (A) 11KT or (B) T production in rainbow darter caught at each site. For both A and B, the horizontal dashed lines indicate the 25% critical

effect size (CES) calculated from the mean of the data from REF2, and the horizontal solid lines indicate the 95% confidence interval calculated from the data from REF2. The vertical gray dashed line indicates the location of the Waterloo WWTP.

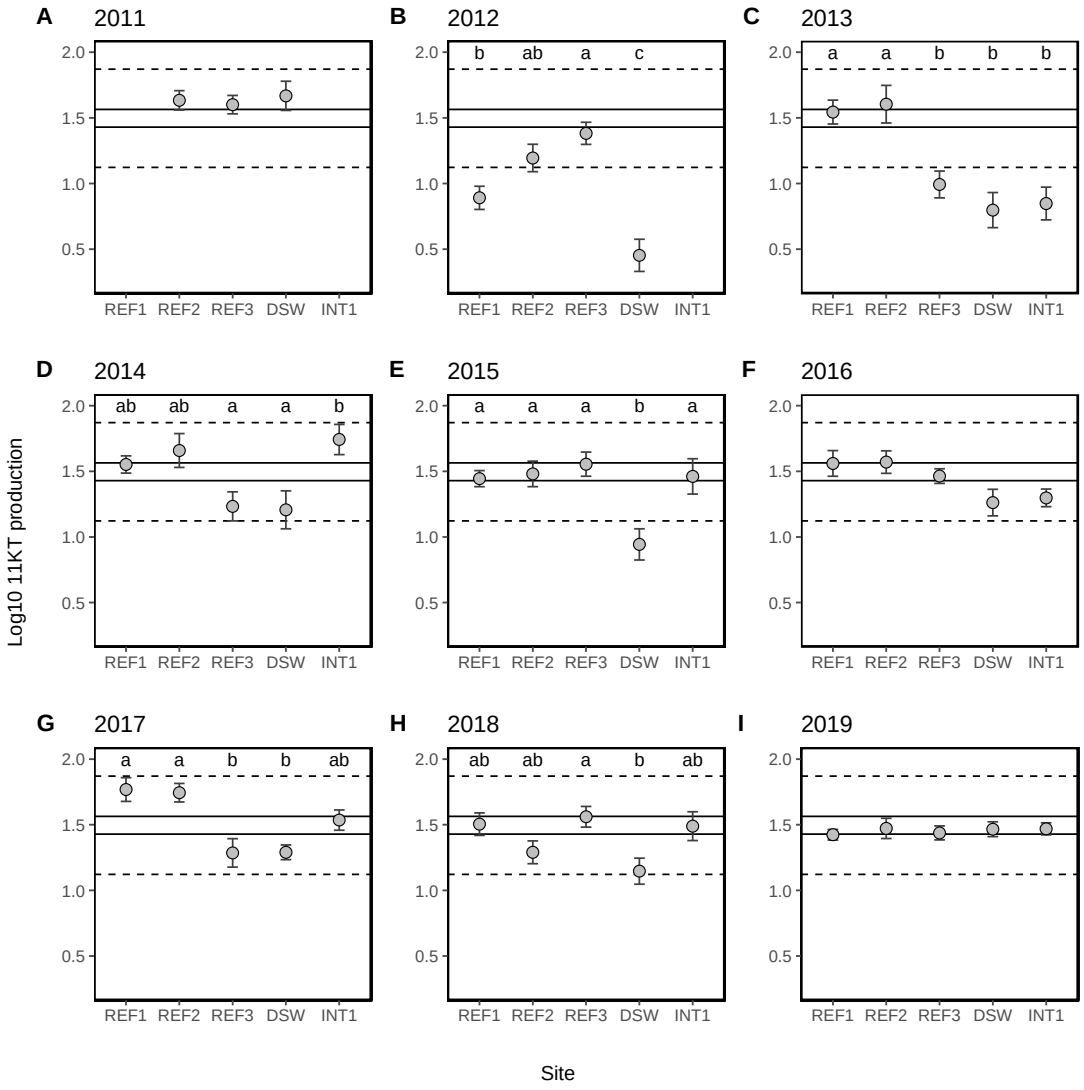

Figure S1.10 Log10 transformed 11KT production (mean ± SE) among sites within each year. Dissimilar letters indicate a statistical difference between sites. The dashed line indicates the 25% critical effect size (CES) calculated from the mean of the data from REF2 (from all years combined), and the solid line indicates the 95% confidence interval calculated from the data from REF2.

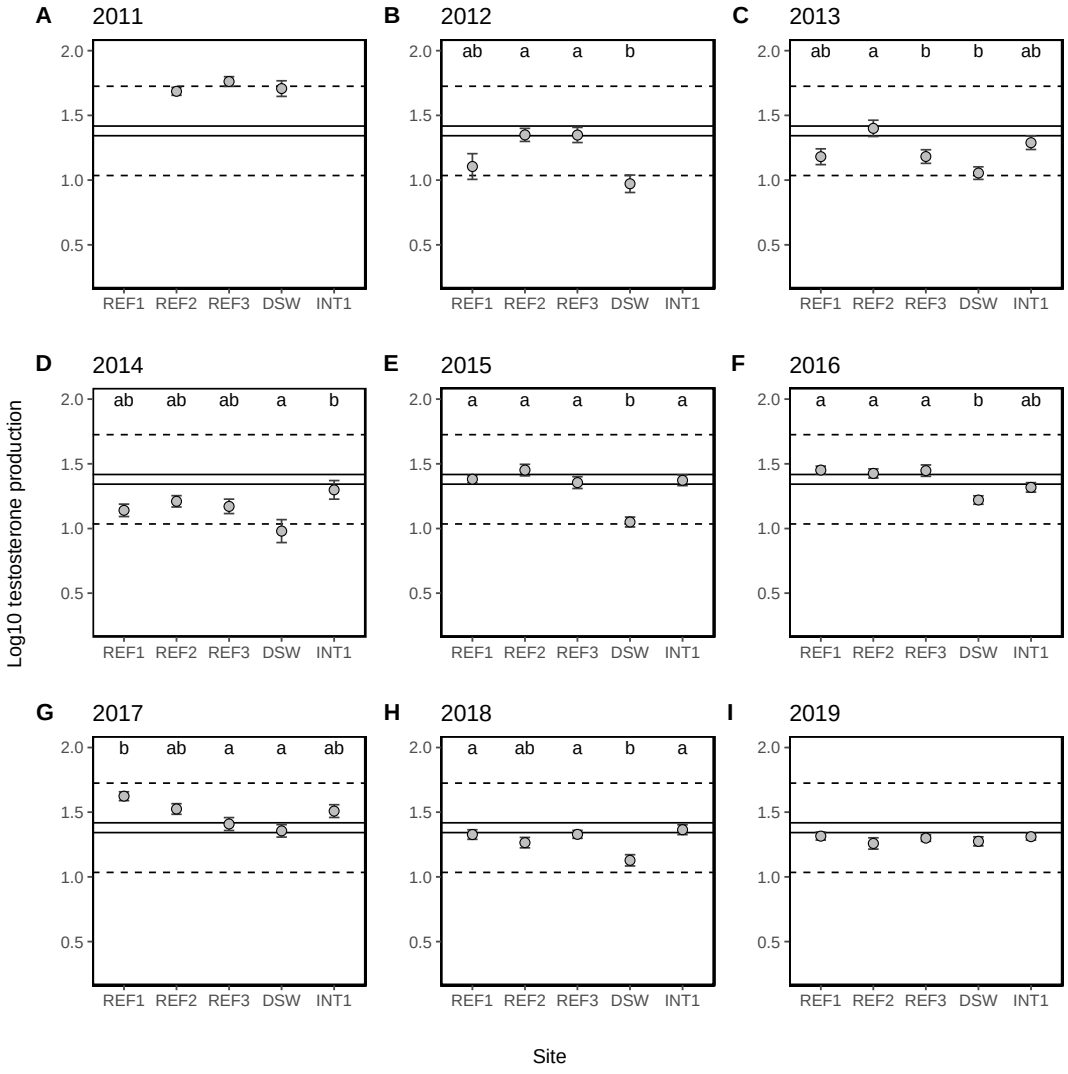

Figure S1.11 Log10 transformed testosterone production (mean ± SE) among sites within each year. Dissimilar letters indicate a statistical difference between sites. The dashed line indicates the 25% critical effect size (CES) calculated from the mean of the data from REF2 (from all years combined), and the solid line indicates the 95% confidence interval calculated from the data from REF2.

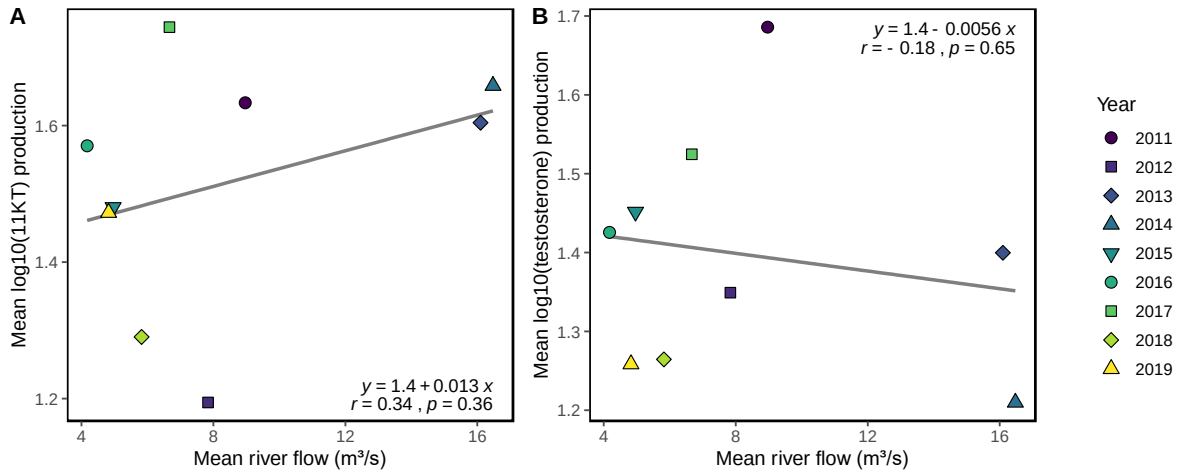

Figure S1.12 Correlation between (A) 11KT or (B) T production in rainbow darter caught at REF2 and mean flow at REF2 from September to October of each year.

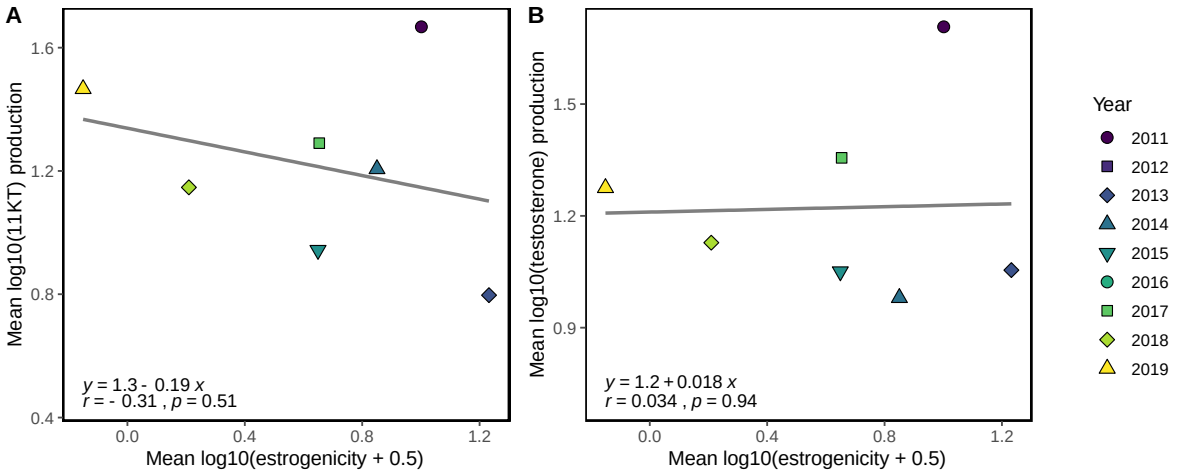

Figure S1.13 Correlation between steroid production of (A) 11KT or (B) T production in rainbow darter caught at DSW and mean estrogenicity (log10 of estrogenicity + 0.5) each year. Estrogenicity sampling was not conducted regularly and therefore the estrogenicity values may not all correspond to a sensitive time period for steroid production in rainbow darter.

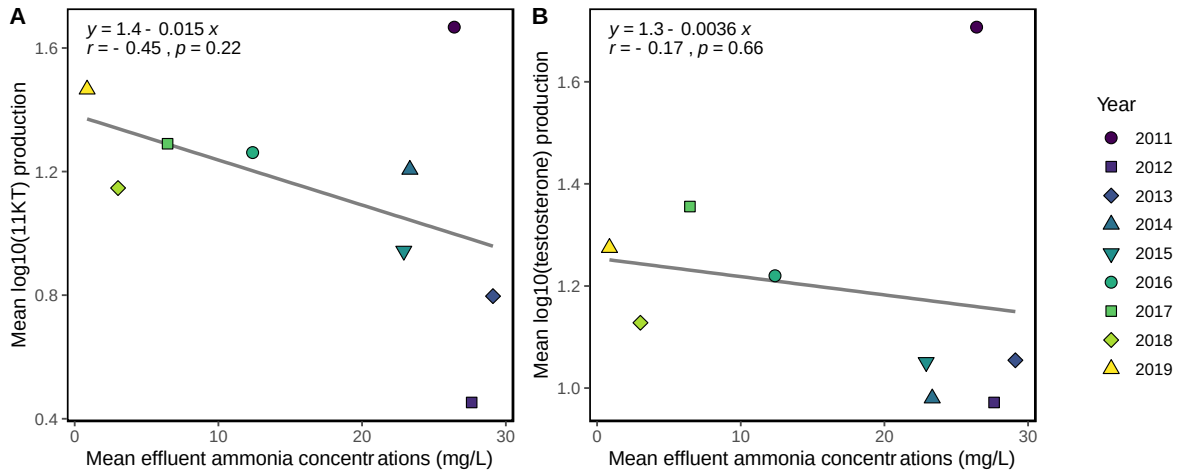

Figure S1.14 Correlation between (A) 11KT or (B) T production in rainbow darter caught at DSW and effluent ammonia concentrations at DSW from September to October of each year.

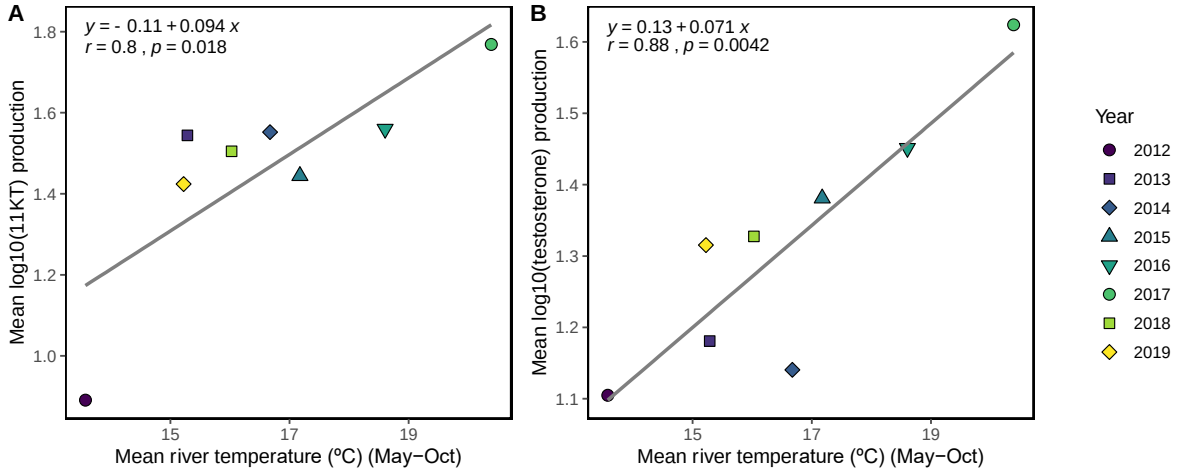

Figure S1.15 Correlation between steroid production of (A) 11KT or (B) T production in rainbow darter caught at REF1 and the mean river water temperature from May to October measured downstream of the Shand Dam (18 km upstream of REF1).

A5 Intersex incidence and severity

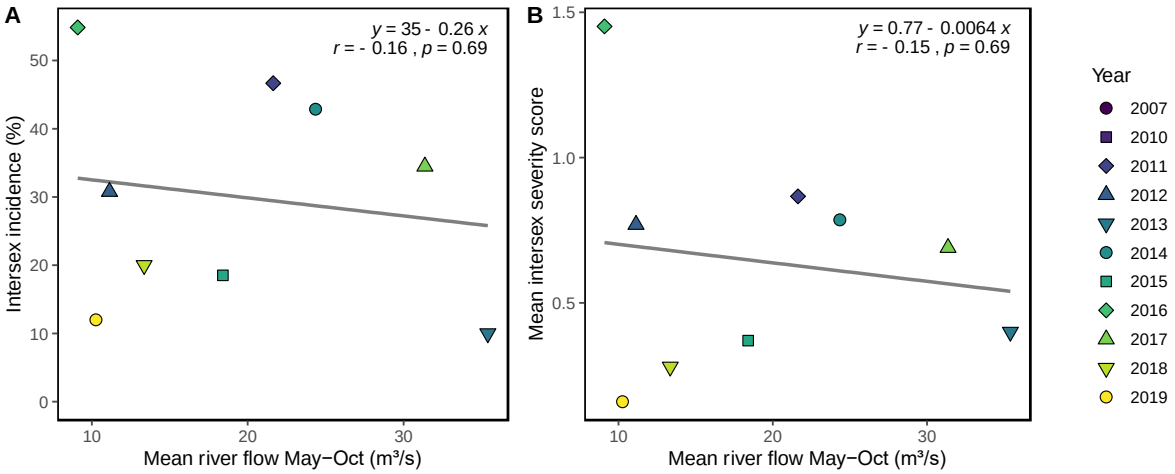

Figure S1.16 Correlation between (A) intersex incidence (%) or (B) intersex severity score in rainbow darter caught at DSW and mean river flows at Bridgeport (0.85 km upstream of the WWTP outfall) from May to October of each year.

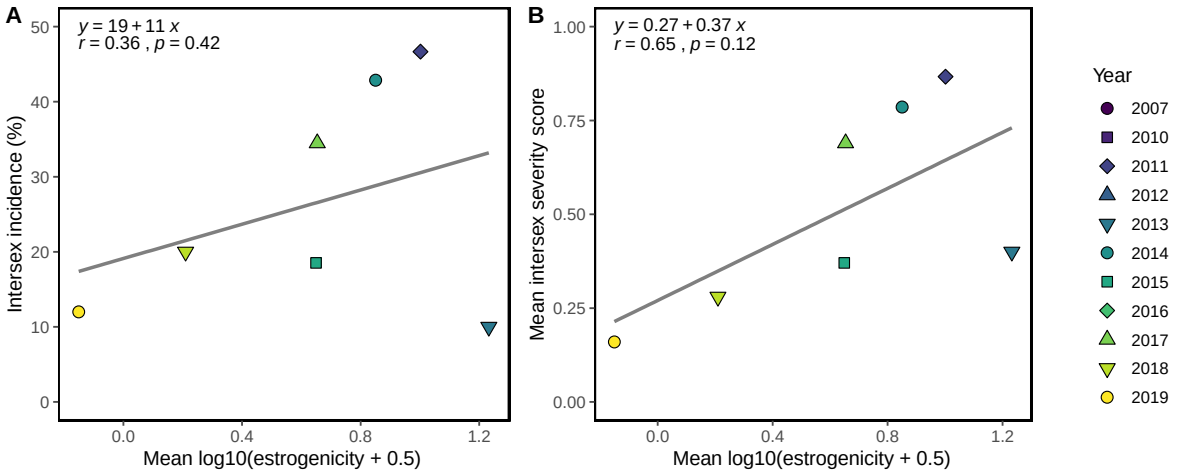

Figure S1.17 Correlation between (A) intersex incidence (%) or (B) mean intersex severity score in fish caught at DSW and mean estrogenicity (expressed as log10(estrogenicity + 0.5) because some measurements = 0), initially measured in ng/L E2eq) in the Waterloo WWTP effluent. Estrogenicity sampling was not conducted regularly and therefore the estrogenicity values may not all correspond to a period of time that was sensitive for gonadal development.

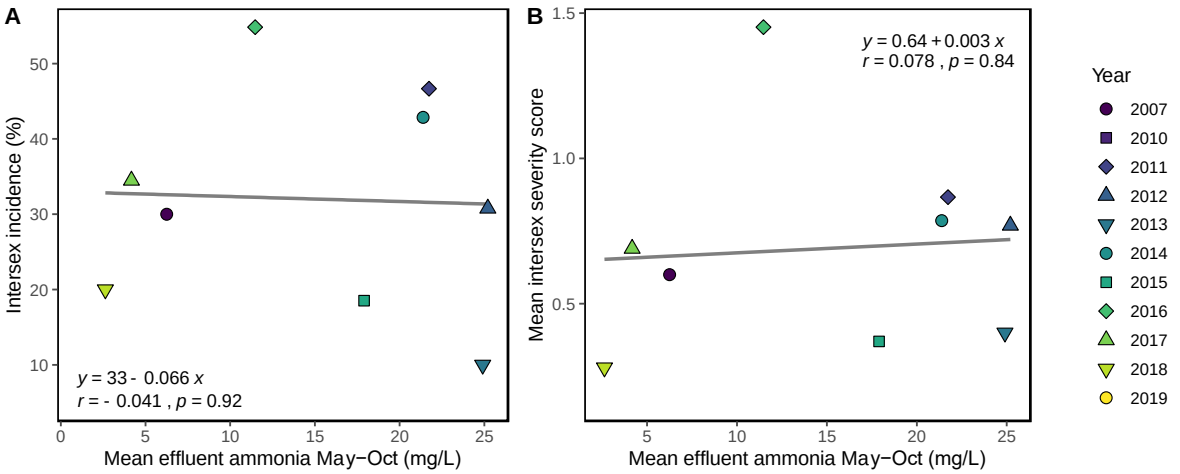

Figure S1.18 Correlation between (A) intersex incidence (%) or (B) intersex severity score in rainbow darter caught at DSW and mean effluent ammonia concentrations from May to October of each year.

Appendix B
Statistical Analyses

B1 Sample sizes

Table S2.1 Sample sizes used in this study for $\delta^{13}C$, $\delta^{15}N$, 11KT, T, and intersex (brackets indicate number of fish with intersex). Superscripts indicate where data has been previously reported.

Year	Site	$\delta^{13}C$ [1]	$\delta^{15}N$ [1]	11KT [2]	T [2]	Intersex(incidence) [3]
2007	REF2	3	3	NA	NA	10 (0)
	REF3	3	3	NA	NA	11 (1)
	DSW	7	7	NA	NA	10 (3)
2008	REF3	6	6	NA	NA	NA
	DSW	6	6	NA	NA	NA
2010	REF1	NA	NA	NA	NA	9 (1)
2011	REF2	12	12	10	10	10 (0)
	REF3	11	11	12	10	10 (1)
	DSW	13	13	10	10	15 (7)
2012	REF1	8	8	9	9	13 (0)
	REF2	8	8	13	12	29 (4)
	REF3	7	8	8	8	12 (1)
	DSW	20	20	12	12	13 (4)
2013	REF1	8	8	9	9	9 (1)
	REF2	8	8	9	10	13 (1)
	REF3	8	8	10	12	16 (2)
	DSW	8	8	9	9	10 (1)
	INT1	8	8	8	8	12 (4)
2014	REF1	7	7	11	11	12 (0)
	REF2	8	7	9	9	13 (1)
	REF3	7	8	10	11	12 (2)
	DSW	8	8	10	10	14 (6)
	INT1	8	8	8	9	19 (4)
2015	REF1	NA	NA	10	11	20 (0)
	REF2	NA	NA	10	9	18 (0)
	REF3	NA	NA	10	10	20 (0)

	DSW	NA	NA	10	10	27 (5)
	INT1	NA	NA	10	10	25 (3)
2016	REF1	8	8	15	15	20 (2)
	REF2	8	8	16	16	22 (1)
	REF3	8	8	14	13	23 (4)
	DSW	8	8	14	14	31 (17)
	INT1	8	8	15	15	24 (7)
2017	REF1	10	10	15	12	25 (3)
	REF2	10	10	15	14	24 (2)
	REF3	10	10	17	15	25 (0)
	DSW	10	10	17	17	29 (10)
	INT1	10	10	15	15	25 (5)
2018	REF1	8	8	20	20	18 (3)
	REF2	8	8	20	20	32 (1)
	REF3	8	8	22	22	22 (0)
	DSW	8	8	20	20	25 (5)
	INT1	8	8	20	20	25 (5)
2019	REF1	NA	NA	20	20	27 (3)
	REF2	NA	NA	20	20	25 (5)
	REF3	NA	NA	20	20	24 (2)
	DSW	NA	NA	20	20	25 (3)
	INT1	NA	NA	20	20	25 (4)

[1] Data from 2007 published in Loomer et al. 2015; data from 2007 to 2014 published in Hicks et al. 2017b

[2] Data from 2011 to 2012 published in Fuzzen et al. 2016; data from 2013 to 2016 published in Marjan et al. 2018

[3] Data from 2007 published in Tetreault et al. 2011; Data from 2010 published in Bahamonde et al. 2014 and Fuzzen et al. 2016; data from 2014 to 2015 published in Hicks et al. 2017a

B2 Stable isotopes

Table S2.2 Summary of one-way ANOVA results (F-ratio with degrees of freedom and *p*-value) for $\delta^{13}C$ and $\delta^{15}N$ among sites for each year studied. Mean ± SE of $\delta^{13}C$ and $\delta^{15}N$ for each site are also reported. Letters are based on Tukey's post-hoc test which was conducted if the one-way ANOVA showed significant differences between sites (dissimilar letters indicate significant differences at alpha = 0.01).

	ANOVA results: $\delta^{13}C$		ANOVA results: $\delta^{15}N$		Tukey's post-hoc results		
Year	F (df)	*p*	F (df)	*p*	Site	Mean ± SE $\delta^{13}C$ (‰)	Mean ± SE $\delta^{15}N$ (‰)
2007	12.7 (2,10)	0.002 *	6.57 (2,10)	0.015	REF2	-28.20 ± 0.18 a	15.27 ± 0.03
					REF3	-27.31 ± 0.19 ab	16.46 ± 0.05
					DSW	-26.18 ± 0.28 b	17.09 ± 0.36
2008	13.5 (1,10)	0.004 *	5.23 (1,10)	0.045	REF3	-27.77 ± 0.20 a	16.24 ± 0.31
					DSW	-26.66 ± 0.23 b	17.30 ± 0.34
2011	22.4 (2,33)	<0.001 *	31.04 (2,33)	<0.001 *	REF2	-28.37 ± 0.07 a	14.77 ± 0.10 a
					REF3	-28.02 ± 0.13 a	16.25 ± 0.05 b
					DSW	-26.79 ± 0.26 b	12.79 ± 0.49 c
2012	27.1 (3,39)	<0.001 *	8.08 (3,40)	<0.001 *	REF1	-27.06 ± 0.13 a	16.46 ± 0.09 a
					REF2	-28.11 ± 0.14 c	15.43 ± 0.05 ab
					REF3	-26.76 ± 0.09 ab	16.55 ± 0.08 a
					DSW	-25.98 ± 0.17 b	14.21 ± 0.45 b
2013	15.3 (4,35)	<0.001 *	62.44 (4,35)	<0.001 *	REF1	-28.18 ± 0.22 a	16.21 ± 0.06 a
					REF2	-28.83 ± 0.18 a	15.38 ± 0.07 a
					REF3	-27.92 ± 0.14 a	16.05 ± 0.06 a
					DSW	-26.55 ± 0.33 b	12.56 ± 0.56 b
					INT1	-28.51 ± 0.20 a	18.71 ± 0.24 c
2014	39.4 (4,33)	<0.001 *	39.49 (4,33)	<0.001 *	REF1	-29.32 ± 0.20 a	16.27 ± 0.07 a
					REF2	-29.80 ± 0.10 a	15.24 ± 0.06 ab
					REF3	-28.55 ± 0.05 b	16.46 ± 0.06 a
					DSW	-27.75 ± 0.13 c	13.87 ± 0.48 b
					INT1	-28.29 ± 0.14 bc	18.46 ± 0.32 c
2016	9.30 (4,35)	<0.001 *	294.14 (4,35)	<0.001 *	REF1	-26.44 ± 0.15 ab	15.23 ± 0.11 a
					REF2	-26.80 ± 0.11 a	14.58 ± 0.10 a
					REF3	-26.61 ± 0.07 a	16.51 ± 0.13 b
					DSW	-25.94 ± 0.19 b	16.66 ± 0.34 b

					INT1	-25.96 ± 0.09 b	23.11 ± 0.20 c
2017	38.8 (4,45)	<0.001 *	87.20 (4,45)	<0.001	REF1	-28.37 ± 0.07 b	14.00 ± 0.14 ab
					REF2	-28.12 ± 0.10 ab	13.76 ± 0.15 a
					REF3	-27.53 ± 0.05 ac	14.97 ± 0.11 b
					DSW	-26.53 ± 0.23 d	16.32 ± 0.42 c
					INT1	-27.06 ± 0.07 cd	18.95 ± 0.16 d
2018	43.5 (4,35)	<0.001 *	128.83 (4,35)	<0.001	REF1	-27.95 ± 0.18 a	15.30 ± 0.08 ab
					REF2	-28.17 ± 0.07 a	14.42 ± 0.16 a
					REF3	-27.77 ± 0.08 a	15.96 ± 0.17 b
					DSW	-26.57 ± 0.11 b	19.94 ± 0.40 c
					INT1	-26.78 ± 0.06 b	19.51 ± 0.18 c

*indicates significance at alpha = 0.01

Table S2.3 Summary of two-way ANOVA results for $\delta^{13}C$ between sites REF3 and DSW and among all years. The interaction between site and year was not significant so the additive model was used. Tukey's post-hoc test was performed and significance was assessed at alpha = 0.01. The mean ± SE $\delta^{13}C$ values are reported, and dissimilar letters indicate significant differences between years or sites.

Two-way ANOVA results	F (df)	*p*
Year	16.38 (8,140)	<0.001 *
Site	119.2 (1,140)	<0.001 *
Tukey's post-hoc test results		Mean ± SE $\delta^{13}C$ (‰)
Year	2007	-26.52 ± 0.26 abc
	2008	-27.21 ± 0.22 a
	2011	-27.35 ± 0.20 a
	2012	-26.19 ± 0.14 bc
	2013	-27.23 ± 0.25 a
	2014	-28.12 ± 0.13 d
	2016	-26.27 ± 0.13 b
	2017	-27.03 ± 0.16 ac
	2018	-27.17 ± 0.17 a
Site	REF3	-27.61 ± 0.08 a
	DSW	-26.49 ± 0.09 b

*indicates significance at alpha = 0.01

Table S2.4 Summary of two-way ANOVA results for $\delta^{15}N$ between sites REF3 and DSW and among all years. The interaction between site and year was significant. Tukey's post-hoc test was performed and significance was assessed at alpha = 0.01. Results are reported for the difference between sites within the same year.

Two-way ANOVA results	F (df)	*p*
Year	1.619 (8,140)	0.12
Site	0.615 (1,140)	0.43
Year x Site	22.000 (8,140)	<0.001 *
Tukey's post-hoc test results		Difference in mean $\delta^{15}N$ (‰) for DSW–REF3 (*p)*
Year	2007	0.63 (0.99)
	2008	1.06 (0.98)
	2011	-3.45 (<0.001 *)
	2012	-2.35 (<0.001 *)
	2013	-3.49 (<0.001 *)
	2014	-2.59 (<0.001 *)
	2016	0.15 (1)
	2017	1.35 (0.47)
	2018	3.98 (<0.001 *)

*indicates significance at alpha = 0.01

B3 *In vitro* steroid production

Table S2.5 Summary of one-way ANOVA results for 11KT and T each year, and mean ± SE steroid production at each site. Dissimilar letters indicate statistical differences following Tukey's post-hoc test assessed at alpha = 0.05 within each year and steroid combination. Dissimilar letters indicate significant differences.

	11KT		T		Steroid production (mean ± SE) with letters indicating significant differences		
Year	F (df)	*p*	F (df)	*p*	Site	log10(11KT)	log10(T)
2011	0.157 (2,29)	0.856	0.763 (2,27)	0.476	REF2	1.63 ± 0.07	1.69 ± 0.03
					REF3	1.60 ± 0.07	1.76 ± 0.04
					DSW	1.67 ± 0.11	1.71 ± 0.06
2012	14.4 (3,38)	<0.001 *	7.64 (3,37)	<0.001 *	REF1	0.89 ± 0.09 b	1.10 ± 0.10 ab
					REF2	1.19 ± 0.10 ab	1.35 ± 0.05 a
					REF3	1.38 ± 0.08 a	1.35 ± 0.06 a
					DSW	0.45 ± 0.12 c	0.97 ± 0.07 b
2013	10.4 (4,40)	<0.001 *	5.22 (4,43)	0.002 *	REF1	1.54 ± 0.09 a	1.18 ± 0.06 ab
					REF2	1.60 ± 0.14 a	1.40 ± 0.06 a
					REF3	0.99 ± 0.10 b	1.18 ± 0.05 b
					DSW	0.80 ± 0.13 b	1.05 ± 0.05 b
					INT1	0.85 ± 0.12 b	1.29 ± 0.05 ab
2014	4.46 (4,43)	0.004 *	3.25 (4,45)	0.020 *	REF1	1.55 ± 0.07 ab	1.14 ± 0.05 ab
					REF2	1.66 ± 0.13 ab	1.21 ± 0.04 ab
					REF3	1.23 ± 0.11 a	1.17 ± 0.06 ab
					DSW	1.21 ± 0.14 a	0.98 ± 0.09 a
					INT1	1.74 ± 0.12 b	1.30 ± 0.07 b
2015	5.61 (4,45)	<0.001 *	15.3 (4,45)	<0.001 *	REF1	1.44 ± 0.06 a	1.38 ± 0.03 a
					REF2	1.48 ± 0.10 a	1.45 ± 0.04 a
					REF3	1.55 ± 0.09 a	1.35 ± 0.05 a
					DSW	0.94 ± 0.12 b	1.05 ± 0.04 b
					INT1	1.46 ± 0.13 a	1.37 ± 0.04 a
2016	3.01 (4,69)	0.024 *†	7.84 (4,68)	<0.001 *	REF1	1.56 ± 0.10	1.45 ± 0.03 a
					REF2	1.57 ± 0.09	1.43 ± 0.04 a
					REF3	1.46 ± 0.06	1.45 ± 0.04 a

					DSW	1.26 ± 0.10	1.22 ± 0.03 b
					INT1	1.30 ± 0.07	1.32 ± 0.04 ab
2017	8.07 (4,74)	<0.001 *	5.02 (4,68)	0.001 *	REF1	1.77 ± 0.09 a	1.62 ± 0.03 b
					REF2	1.74 ± 0.07 a	1.52 ± 0.04 ab
					REF3	1.29 ± 0.11 b	1.41 ± 0.05 a
					DSW	1.29 ± 0.06 b	1.36 ± 0.05 a
					INT1	1.54 ± 0.08 ab	1.51 ± 0.05 ab
2018	3.57 (4,97)	0.009 *	6.05 (4,97)	<0.001 *	REF1	1.50 ± 0.09 ab	1.33 ± 0.04 a
					REF2	1.29 ± 0.09 ab	1.26 ± 0.04 ab
					REF3	1.56 ± 0.08 a	1.33 ± 0.03 a
					DSW	1.15 ± 0.10 b	1.13 ± 0.04 b
					INT1	1.49 ± 0.11 ab	1.36 ± 0.04 a
2019	0.150 (4,95)	0.962	0.557 (4,95)	0.695	REF1	1.42 ± 0.04	1.32 ± 0.03
					REF2	1.47 ± 0.08	1.26 ± 0.04
					REF3	1.44 ± 0.05	1.30 ± 0.02
					DSW	1.47 ± 0.06	1.27 ± 0.04
					INT1	1.47 ± 0.05	1.31 ± 0.03

*Indicates significance at alpha = 0.05

†ANOVA was significant but no pairwise comparisons were significant

Table S2.6 Summary of one-way ANOVA results for steroid production at DSW among years (using data that was logged and normalized to REF2). Mean ± SE normalized production are reported with Tukey's post-hoc test assessed at alpha = 0.05. Dissimilar letters indicate significant differences.

	11KT		T		Logged and normalized to REF2 steroid production (mean ± SE) with letters indicating significance		
	F (df)	*p*	F (df)	*p*	Year	11KT	T
Year	8.88 (8,113)	<0.001 *	8.16 (8,113)	<0.001 *	2011	1.02 ± 0.07 d	1.01 ± 0.04 cd
					2012	0.38 ± 0.10 a	0.72 ± 0.05 a
					2013	0.50 ± 0.08 ab	0.75 ± 0.03 ab
					2014	0.73 ± 0.09 bcd	0.81 ± 0.07 ab
					2015	0.64 ± 0.08 abc	0.72 ± 0.03 ab
					2016	0.80 ± 0.06 bcd	0.86 ± 0.02 abc
					2017	0.74 ± 0.03 bcd	0.89 ± 0.03 bcd
					2018	0.89 ± 0.08 cd	0.89 ± 0.03 bcd
					2019	1.00 ± 0.04 d	1.01 ± 0.03 d

*Indicates significance at alpha = 0.05

B4 Intersex incidence and severity

Table S2.7 Fisher's exact test results for intersex incidence data and Kruskal–Wallis test results (test statistic = H) for intersex severity score among years within each site. Only the significant pairwise comparisons ($p < 0.05$) are listed for each post-hoc test. All other pairwise comparisons were not significantly different.

	Incidence: Fisher's Exact Test		Severity Score: Kruskal–Wallis Test		
Site	Fisher's Exact Test *p*	Significant pairwise comparisons (*p*)	H (df)	*p*	Significant pairwise comparisons (*p*)
REF1	0.534		6.51 (8)	0.590	
REF2	0.361		11.47 (9)	0.245	
REF3	0.114		18.78 (9)	0.291	
DSW	0.013 *	†	23.30 (9)	0.006 *	2015–2016 (0.040) 2016–2018 (0.047) 2016–2019 (0.006)
INT1	0.714		4.34 (6)	0.630	

*Indicates significance at alpha = 0.05

†No significant differences after adjusting for multiple pairwise comparisons

Table S2.8 Fisher's exact test results for intersex incidence data and Kruskal–Wallis test results (test statistic = H) for intersex severity score among sites within each year. Only the significant pairwise comparisons ($p < 0.05$) are listed for each post-hoc test. All other pairwise comparisons were not significantly different.

	Incidence: Fisher's Exact Test		Severity Score: Kruskal–Wallis Test		
Year	*p*	Significant pairwise comparisons (*p*)	H (df)	*p*	Significant pairwise comparisons (*p*)
2007	0.184		3.73 (2)	0.155	
2011	0.016 *	†	7.87 (2)	0.020 *	REF2–DSW (0.027)
2012	0.144		6.15 (3)	0.104	
2013	0.513		4.31 (4)	0.366	
2014	0.056		9.75 (4)	0.045 *	REF1–DSW (0.043)
2015	0.028 *	†	10.44 (4)	0.035 *	†
2016	<0.001 *	REF1–DSW (0.011) REF2–DSW (0.002)	25.20 (4)	<0.001 *	REF1–DSW (0.001) REF2–DSW (<0.001) REF3–DSW (0.008)
2017	0.005 *	REF2–DSW (0.001)	14.85 (4)	0.005 *	REF2–DSW (0.004)
2018	0.031 *	†	9.14 (4)	0.057	
2019	0.812		1.66 (4)	0.798	

*Indicates significance at alpha = 0.05

†No significant differences after adjusting for multiple pairwise comparisons